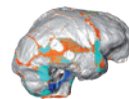


THIS WEEK

EDITORIALS

GETTING BETTER *Nature* tackles sloppy science with your help **p.398**

WORLD VIEW The danger signs are there for new flu outbreak **p.399**



ORIGINS Hobbit humans had larger brains, model study shows **p.400**

The fight against bird flu

China's well-handled response to outbreaks of H7N9 avian influenza belies the country's bad reputation from its past dealings with disease. But there are still improvements to be made.

China deserves credit for its rapid response to the outbreaks of H7N9 avian influenza, and its early openness in the reporting and sharing of data.

A bad reputation is difficult to shake. A decade ago, China failed to report early cases of severe acute respiratory syndrome (SARS) and fumbled its initial response to the threat. Today, some commentators view its reaction to H7N9 with mistrust. But from all the evidence so far, China's response to the virus, which had caused 104 confirmed human cases and 21 deaths as *Nature* went to press, is next to exemplary.

China reported the H7N9 outbreak to the World Health Organization (WHO) on 31 March, just six weeks after the first known person fell ill. On the same day, it published the genomic sequences of viruses from the three human cases then identified on the database of the Global Initiative on Sharing Avian Influenza Data (GISAID). It has also shared all the sequences with the WHO, and live virus with the WHO and other laboratories. This has allowed scientists to identify the virus's mutations, trace its origins and develop crucial diagnostic tests. China continues to report new cases daily, and its media discusses H7N9 fairly openly. Chinese and other researchers have quickly published detailed analyses of the virus in journals (R. Gao *N. Engl. J. Med.* <http://doi.org/k7r>; 2013). Chinese President Xi Jinping added political clout last week when he called for an effective response, and said that the government must ensure the release of accurate information about the outbreaks.

China's response to the epidemic has also been brisk. Diagnostic tests have been distributed to hospitals and research labs across the country. The response, spearheaded by the Chinese Center for Disease Control and Prevention in Beijing, has united clinicians, virologists, and epidemiologists. Live-bird markets at which H7N9 has been found have been shut down, and birds culled. The agriculture ministry has tested tens of thousands of birds and other animals for the virus, to try to pin down the sources of human infections and explain their occurrence in cities hundreds of kilometres apart — no mean task given that China has some 6 billion domestic fowl and half a billion pigs, which can also carry the virus. So far, however, apart from birds at the live markets, the sources of infection remain elusive. To help track them down, and to collaborate in efforts to control H7N9, China has invited a team of WHO scientists and international flu experts to the country. They arrived last week, and are expected to report their preliminary conclusions this week.

Yet suspicions linger. Some critics have questioned, for example, the time between the first person falling ill on 19 February and China's first announcement about the virus, and have asked whether the announcement was deliberately delayed. This is unfair. With just a handful of severe pneumonia cases caused by the virus by mid-March, it is impressive that China realized as quickly as it did that something was amiss. It took the United States, which has one of the world's most advanced disease-surveillance systems, an almost identical amount of time to identify a novel H3N2 swine virus that caused serious illness in a child in 2011.

China has made a good start, but it is crucial for the country to

continue its openness over the H7N9 outbreaks. In particular, it must promptly report any evidence of human-to-human spread. There are also areas for improvement: data made public on human cases are often limited to basic facts such as age, sex, date of onset of illness and location. Epidemiologists also need more detailed data, including possible exposures to infection and underlying medical conditions. Case reports should be published in full in journals or online as quickly as possible.

It is also important that sequences from as many cases as possible are submitted to publicly accessible databases, because sequence data are important in tracking evolutionary changes such as new mutations that could allow the virus to spread between humans more easily. They can also provide clues to the source of infection (see page 399).

"China has made a good start, but it is crucial for the country to continue its openness over the H7N9 outbreaks."

Even as the Chinese authorities are being open and transparent on H7N9, some scientists are hoarding epidemiological and other data, because of intense competition to be the first to publish. Competition can be healthy, but in the face of a virus that has the potential to cause a pandemic, researchers have a duty above all else to share important data. Journals must be ready and willing, as in any public-health emergency, to fast-track peer review of H7N9 papers, and not let rapid publication of preprints stand in the way of considering papers for publication. Meanwhile, observers should continue to scrutinize China's response to H7N9, but they should also give credit where credit is due. It is time to recognize that China has changed. ■

Across the divide

Diagnostic boundaries separating mental disorders hamper effective treatments.

Scientists who attended the 2009 Winter Workshop in Psychoses in Barcelona, Spain, may not have realized it at the time, but they were part of a revolution. In previous years, organizers named the event the Winter Workshop on Schizophrenia and Bipolar Disorders. It was one of the few conferences at which those who studied schizophrenia and those who worked on bipolar illnesses would meet.

As Nick Craddock, a psychiatrist who studies both conditions at Cardiff University, UK, says in a News Feature on page 416, a merger of these two distinct groups — even in semantic terms — would have been unthinkable until very recently. Psychiatrists diagnose

schizophrenia and bipolar disorder as two separate conditions. This separation is respected by drug companies, regulators, research funders, journals and bench researchers. Add that lot up, and you get a fundamental problem with psychiatry.

Next month, the American Psychiatric Association will release the long-awaited fifth version of its *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), which lists mental illnesses and their symptoms. Work on preparing the DSM-5 has been clouded in controversy, and the arguments over which conditions should have been included and which left out will rumble on for some time.

The more fundamental problem, as the News Feature explores, is growing doubt about the way the DSM-5 classifies mental disorders. Psychiatrists have long known that the illnesses of patients they see in the clinic cannot be broken down into discrete groups in the way that is taught at medical school. Symptoms overlap and flow across diagnostic boundaries. Patients can show the signs of two or three disorders at the same time. Treatments are inconsistent. Outcomes are unpredictable.

Science was supposed to come to the rescue. Genetics and neuro-imaging studies would, all involved hoped, reveal biological signatures unique to each disorder, which could be used to provide consistent

and reliable diagnoses. Instead, it seems the opposite is true. The more scientists look for biomarkers for specific mental disorders, the harder the task becomes. Scans of the DNA and brain function of patients show the same stubborn refusal to group by disease type. Genetic risk factors and dysfunction in brain regions are shared across disorders.

“Patients’ illnesses cannot be broken down into discrete groups in the way that is taught at medical school.”

Psychiatrists joke that their patients have not read the textbooks. The reality is serious and more troubling — the textbook is wrong.

The American Psychiatric Association routinely points out that its DSM disease categories are intended only as diagnostic tools. It does not claim that they mark genuine biological boundaries. But the system is set up as if they do. That might explain why biomarkers and new drugs for mental illness remain elusive. The system should change. Funders and journals must encourage work that cuts across the boundaries. Researchers should be encouraged to investigate the causes of mental illness from the bottom up, as the US National Institute of Mental Health is doing. The brain is complicated enough. Why investigate its problems with one hand tied behind our backs? ■

ANNOUNCEMENT

Reducing our irreproducibility

Over the past year, *Nature* has published a string of articles that highlight failures in the reliability and reproducibility of published research (collected and freely available at go.nature.com/huhbyr). The problems arise in laboratories, but journals such as this one compound them when they fail to exert sufficient scrutiny over the results that they publish, and when they do not publish enough information for other researchers to assess results properly.

From next month, *Nature* and the Nature research journals will introduce editorial measures to address the problem by improving the consistency and quality of reporting in life-sciences articles. To ease the interpretation and improve the reliability of published results we will more systematically ensure that key methodological details are reported, and we will give more space to methods sections. We will examine statistics more closely and encourage authors to be transparent, for example by including their raw data.

Central to this initiative is a checklist intended to prompt authors to disclose technical and statistical information in their submissions, and to encourage referees to consider aspects important for research reproducibility (go.nature.com/oloqip). It was developed after discussions with researchers on the problems that lead to irreproducibility, including workshops organized last year by US National Institutes of Health (NIH) institutes. It also draws on published concerns about reporting standards (or the lack of them) and the collective experience of editors at Nature journals.

The checklist is not exhaustive. It focuses on a few experimental and analytical design elements that are crucial for the interpretation of research results but are often reported incompletely. For example, authors will need to describe methodological parameters that can introduce bias or influence robustness, and provide precise characterization of key reagents that may be subject to biological variability, such as cell lines and antibodies. The checklist also consolidates existing policies about data deposition and presentation.

We will also demand more precise descriptions of statistics, and

we will commission statisticians as consultants on certain papers, at the editor's discretion and at the referees' suggestion.

We recognize that there is no single way to conduct an experimental study. Exploratory investigations cannot be done with the same level of statistical rigour as hypothesis-testing studies. Few academic laboratories have the means to perform the level of validation required, for example, to translate a finding from the laboratory to the clinic. However, that should not stand in the way of a full report of how a study was designed, conducted and analysed that will allow reviewers and readers to adequately interpret and build on the results.

To allow authors to describe their experimental design and methods in as much detail as necessary, the participating journals, including *Nature*, will abolish space restrictions on the methods section.

To further increase transparency, we will encourage authors to provide tables of the data behind graphs and figures. This builds on our established data-deposition policy for specific experiments and large data sets. The source data will be made available directly from the figure legend, for easy access. We continue to encourage authors to share detailed methods and reagent descriptions by depositing protocols in Protocol Exchange (www.nature.com/protocolexchange), an open resource linked from the primary paper.

Renewed attention to reporting and transparency is a small step. Much bigger underlying issues contribute to the problem, and are beyond the reach of journals alone. Too few biologists receive adequate training in statistics and other quantitative aspects of their subject. Mentoring of young scientists on matters of rigour and transparency is inconsistent at best. In academia, the ever increasing pressures to publish and chase funds provide little incentive to pursue studies and publish results that contradict or confirm previous papers. Those who document the validity or irreproducibility of a published piece of work seldom get a welcome from journals and funders, even as money and effort are wasted on false assumptions.

Tackling these issues is a long-term endeavour that will require the commitment of funders, institutions, researchers and publishers. It is encouraging that NIH institutes have led community discussions on this topic and are considering their own recommendations. We urge others to take note of these and of our initiatives, and do whatever they can to improve research reproducibility. ■



H7N9 is a virus worth worrying about

Warnings about the emergence of another influenza virus may elicit scepticism, but we should not be complacent, cautions Peter Horby.

Once again an animal influenza A virus has crossed the species barrier to cause an appreciable number of human cases. Now, two months after the first known human infections with the H7N9 virus, the question is: which of the paths set by previous emerging influenza viruses will it follow?

One predecessor, H5N1, generated alarm owing to its high pathogenicity in humans. It has proved to be a tenacious adversary, remaining endemic in poultry across large parts of Asia, but thankfully it has not adapted to humans and person-to-person transmission remains rare. A second, H7N7, caused a number of mostly mild human infections in the Netherlands in 2003, with some evidence of limited person-to-person spread, but extensive poultry culling controlled it. A third, the H1N1 swine influenza virus that emerged in 2009, successfully adapted to humans and caused a pandemic.

So will H7N9 prove to be controllable? Will it remain entrenched in animals? Or will it, like the H1N1 virus, stably adapt to humans and cause a pandemic? The fine line between foresight and alarmism can only be drawn in retrospect. Nevertheless, my colleagues and I consider that H7N9 has many of the traits that make a new flu virus worrisome.

The H7N9 haemagglutinin protein — which binds to target cells — resembles those of other avian flu viruses that cause only mild disease in birds. This means that the virus is likely to spread silently in domestic and probably wild birds. Human infections are therefore the sentinel events, and the numbers and geographic extent of human cases — all of them so far in China — suggest that a hidden epidemic in other animals is well under way.

The small number of poultry in which H7N9 has so far been detected is rather puzzling, as are the 20% of people infected with the virus who have not reported exposure to poultry. Nevertheless, domestic birds are likely to be the main source of human infections. And the animal epidemic is likely to spread farther, with large suppliers distributing poultry across China. Flying wild birds are another possible mode of spread. Given that the virus probably does not cause severe disease in birds, and the uncertainty surrounding the animal source, containing the animal epidemic poses an enormous challenge.

So far, extensive monitoring of contacts has not found evidence that the virus has spread efficiently between people. Limited human-to-human transmission may have occurred but, as we saw with H5N1 and H7N7, this does not necessarily represent the early stages of a trajectory towards full human adaptation. However, H7N9 viruses isolated from patients possess some genetic signatures that are associated with effective replication and transmission, and with high virulence in mammals. The regions of China

where H7N9 seems to be circulating have large populations of pigs as well as humans, providing opportunities for further adaptation to mammals and for re-assortment with human- or pig-adapted viruses.

The clinical epidemiology of H7N9 cases has some similarities to human seasonal influenza. Unlike the H7N7 cases in 2003, which usually took the form of conjunctivitis, the H7N9 infections so far detected have caused respiratory illness, with cases in all ages but being most severe in the elderly and people with underlying illnesses. However, the fact that the average age of people infected is high — around 60 years — and that most reported infections have been severe suggests that the virus is not yet well adapted to humans. Only further clinical and epidemiological data will reveal the full spectrum of infection and severity.

Standardized collection and sharing of clinical data would aid risk assessment and treatment. A clinical protocol and case-record and informed-consent forms developed by the International Severe Acute Respiratory and Emerging Infection Consortium and the World Health Organization are available online (see go.nature.com/fpsiog).

If H7N9 were to stably adapt to humans, it would probably meet with little or no human immunity. Detecting and tracking a partially human-adapted H7N9 virus in a city as vast as Shanghai or Beijing would be difficult; tracking a fully adapted virus would be impossible. And it could easily spread nationally and internationally. Eastern China is now one of the most 'connected' population centres in the world. Seventy per cent of the global popula-

tion outside China lives within two hours of an airport linked to the outbreak regions by a direct flight or a single connection (see go.nature.com/tvfev8). Travel restrictions or border screening will not contain pandemic influenza for long.

If there was an overreaction to H1N1, we should not compound the error by under-reacting to H7N9. Hopefully H7N9 will remain an animal virus, and maybe the fact that it has circulated for at least two months without stably adapting to humans indicates that the species barrier is too great for it; but maybe not. The first human case of H7N9 outside mainland China is perhaps only a matter of time. Then the public-health and clinical community will need to assess, carefully and quickly, whether it represents a single imported case of animal-to-human transmission, an animal epidemic that has spread abroad, or the international spread of a partially or fully human-adapted virus. ■

Peter Horby is based at the Oxford University Clinical Research Unit, Wellcome Trust Programme Vietnam and the Singapore Infectious Disease Initiative. This article reflects the views and expertise of many colleagues, who are listed at go.nature.com/lskoqj. e-mail: phorby@oucru.org

THE FIRST
HUMAN CASE
OF H7N9 OUTSIDE
MAINLAND CHINA IS
PERHAPS
ONLY A MATTER OF
TIME.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/lskoqj

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

NEUROSCIENCE

Stimulating depression away

Patients with treatment-resistant depression showed rapid improvement after electrodes were inserted at a site in the medial forebrain — a region associated with motivation and reward.

Of the seven patients who received deep brain stimulation, Volker Coenen at University Hospital Freiburg, Germany, and his colleagues report that six responded — measured by a common scale of depression — within days. This response is much faster than the many weeks required for an antidepressant effect in other pilot studies in which researchers targeted other sites in the same brain region and used a higher current. However, the authors say that their results are preliminary and need to be confirmed in larger, controlled studies.

Biol. Psychiatry <http://dx.doi.org/10.1016/j.biopsych.2013.01.034> (2013)

MICROBIOLOGY

Dogs and owners share microbes

Humans are colonized by the same types of microbe as the people and the pets they live with.

Rob Knight at the University of Colorado Boulder and his team used DNA sequencing to analyse the microbes colonizing the skin, guts and mouths of 159 people and 36 dogs, living in 60 households.

Humans tended to have similar microbial communities — particularly on the skin — to their spouses and children. Adult dog-owners also had more skin microbes in common with their dogs than with other dogs. However,

microbes in the mouths and guts of canines differed from those of their owners. Shared skin microbiota might help to explain why dog ownership is associated with reduced allergy rates in children, the researchers say.

eLIFE 2, e00458 (2013)

BIOMATERIALS

Worm-inspired adhesive

Whether tissue is wet or dry, a new bandage will stick to it — like an intestinal parasite.

Jeffrey Karp at Brigham and Women's Hospital in Boston,

Massachusetts, and his team have designed a gripping material that steals the sticky secrets of the spiny-headed worm *Pomphorhynchus laevis*. The parasite pierces its fish host with a proboscis that then swells up to lock into place.

The researchers' adhesive is made up of spikes coated with a super-absorbent plastic. When the spikes come into contact with water in tissue, they swell and fasten to the tissue. The removable bandage adhered tightly to pig skin and intestine, and was more than three times as adhesive as surgical staples for affixing skin grafts.

Nature Commun. 4, 1702 (2013)



RYAN TAYLOR

ANIMAL BEHAVIOUR

Babies of stressed squirrels grow faster

Social stress alters hormone levels in red-squirrel mothers and leads to faster-growing pups.

In a 22-year study, Ben Dantzer, now at the University of Cambridge, UK, and his team found that, in densely populated red-squirrel (*Tamiasciurus hudsonicus*; pictured) communities, females that had faster-growing pups saw more of them survive their first winter. The researchers simulated crowded conditions by playing recordings of squirrels'

territorial cries. Mothers living in dense groups or exposed to the cries had higher levels of breakdown products from the stress hormone cortisol in their faeces. Pups of squirrels that heard the recordings grew faster than pups of females that heard bird noises. Feeding pregnant squirrels cortisol also boosted the growth rate of their pups, by 41%.

Science <http://dx.doi.org/10.1126/science.1235765> (2013)

NEUROSCIENCE

Autism gene alters endocannabinoids

Certain gene mutations associated with autism interfere with nervous-system signals that activate the same pathways as cannabis.

Some people with autism have a mutation or deletion in the neuronal gene *neuroligin-3*. But, although mice carrying a gene with the human mutation show autism-like behaviours, mice that lack the gene do not. Csaba Földy, Robert Malenka and Thomas Südhof at Stanford University

in California have found that mutation, as well as deletion, of the gene changes how certain groups of neurons in the brain transmit signals. Signalling of a neuronal receptor that responds to cannabis and to endocannabinoids (which are made by the brain) is impaired in mice in which neuroligin-3 is mutated or missing. This suggests that disrupted endocannabinoid signalling contributes to autism, a mechanism that could suggest new strategies for treatment.

Neuron <http://dx.doi.org/10.1016/j.neuron.2013.02.036> (2013)

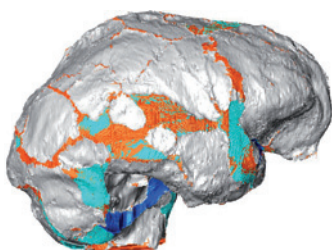
PALAEOONTOLOGY

'Hobbit' brains not so small

New estimates of brain size for *Homo floresiensis* make it more feasible that the diminutive hominid descended from *Homo erectus*.

The origins of *H. floresiensis* have been intensely debated in the decade since the roughly 18,000-year-old fossils of the 1-metre-tall hominid were discovered on the island of Flores in eastern Indonesia. Yousuke Kaifu at the University of Tokyo and his colleagues used replicas of an *H. floresiensis* skull and high-resolution computed-tomography scans to make models (pictured) of the hominid's brain. Their calculation of 426 cubic centimetres — roughly one-third the volume of a human brain, and the most accurate estimate so far — is slightly bigger than previous estimates. Just big enough, the authors say, that it is mechanistically possible that *H. erectus* underwent extreme dwarfism on an isolated island.

Proc. R. Soc. B 280, 20130338 (2013)



ALMA/J. HODGE/A. WEISS

DAISUKE KUBO

ECOLOGY

Seeds travel on unpaved roads

Dirt roads could be providing important corridors for seed distribution.

Alberto Suárez-Esteban and his colleagues at the Doñana Biological Station in Seville, Spain, collected animal faeces from 66 kilometres of man-made breaks in vegetation, such as firebreaks and dirt roads, as well as adjacent scrubland, in Doñana National Park in southwest Spain. The researchers identified and counted the seeds contained in 615 faecal samples from rabbits, carnivores and ungulates such as deer.

Carnivores and rabbits preferred to defecate on tracks, dispersing up to 124 times as many viable seeds along the tracks as in the scrub. Although ungulates avoided defecating along the tracks, their faeces also contained considerably fewer viable seeds.

The authors suggest that such human disruptions could have an overlooked role in plant conservation by helping animals to spread seeds between isolated plant populations, but they could also provide routes for invading species.

J. Appl. Ecol. <http://dx.doi.org/10.1111/1365-2664.12080> (2013)

CLIMATE CHANGE

Evolution in acidic oceans

An increase in ocean acidity could drive substantial genetic change in sea urchins in just one generation.

Melissa Pespeni at Hopkins Marine Station in Pacific Grove, California, and her colleagues housed developing purple sea urchins (*Strongylocentrotus purpuratus*) under current acidity levels and the higher levels that are expected from increasing amounts of carbon dioxide in the

COMMUNITY CHOICE

The most viewed papers in science

MATERIALS

Nanospheres make clever membranes

HIGHLY READ
on www.acs.org
in March

Spheres of silica coated in gold have been made into membranes whose permeability can be engineered.

Ilya Zharov and Patricia Ignacio-de Leon at the University of Utah in Salt Lake City created nanospheres that self-assemble into arrays, and can then be heated to make inorganic membranes. By coating the silica spheres with gold, the duo were able to attach a variety of chemical groups to the spheres. Surface modifications affected how various molecules passed through the membranes, a process that could be further controlled by changes in pH. Such materials could have applications in chemical separations, catalysts and sensors, the authors say. **Langmuir** 29, 3749–3756 (2013)

atmosphere. The authors measured changes in the frequency of 19,493 genetic variants as fertilized eggs grew into swimming and feeding larvae. Although conditions of high acidity had little effect on the growth of the animals, major shifts occurred in genes that code for 40 classes of proteins. These changes were concentrated in genes related to the construction of urchins' shells and how the organisms regulate metabolism and pH.

Increased acidity could be selecting for genetic variants that improve survival under these conditions, the researchers suggest.

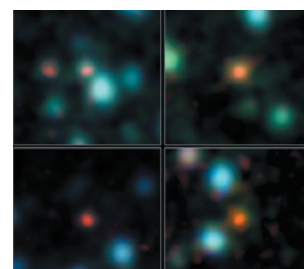
Proc. Natl Acad. Sci. USA <http://dx.doi.org/10.1073/pnas.1220673110> (2013)

ASTRONOMY

Dusty galaxies come into view

Astronomers have made their first statistically reliable survey of one kind of star-forming galaxy in the early Universe.

Knowledge of these distant objects is important for our understanding of these galaxies' formation and evolution, but enshrouding dust usually obscures their



details — making them hard to identify with telescopes that collect radio waves or visible light. Jacqueline Hodge at the Max Planck Institute for Astronomy in Heidelberg, Germany, and her colleagues used the Atacama Large Millimeter/submillimeter Array (ALMA) in Chile to penetrate the dust veil by looking for emissions at submillimetre wavelengths of light — a length between infrared and radio waves.

The scientists' observation of 126 previously unresolved galaxies in the southern constellation Fornax brought blurry objects into sharper focus (pictured). At least one-third, and possibly up to one-half, of them turned out to be multiple galaxies.

Astrophys. J. 768, 91 (2013)

NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch

SEVEN DAYS

The news in brief

EVENTS

Toxic letters

A Mississippi man suspected of sending letters laced with the deadly toxin ricin to US President Barack Obama and Republican senator Roger Wicker (Mississippi) was charged with threatening injury and death by the Federal Bureau of Investigation on 18 April. The letters were intercepted the previous day at postal screening facilities, and lab tests confirmed the presence of ricin, which is 1,000 times more toxic than cyanide. The toxin is most lethal when inhaled or injected; there is no antidote, but symptoms can be treated. The suspect, Paul Kevin Curtis, of Corinth, faces up to 15 years in prison if convicted. See go.nature.com/xgnin4 for more.

Rocket launch

The Antares rocket built by Orbital Sciences of Dulles, Virginia, successfully completed its maiden test flight on 21 April. The rocket is the first vehicle to take off from NASA's new launch pad at the Wallops Flight Facility in Virginia. The flight puts NASA one step closer to having two US cargo carriers available to resupply the International Space Station. "It looks like it performed flawlessly throughout the

NUMBER CRUNCH

\$674 bn

The amount spent last year finding and developing new fossil-fuel reserves, according to Carbon Tracker, even though burning them would cause a catastrophic rise in global temperatures.



HERITAGE AUCTIONS

Return of the dinosaur

Mongolian researchers have finally got their hands on the *Tarbosaurus bataar* fossil (pictured) that was illegally smuggled out of Mongolia and came to auction in the United States last year. On 16 April, the team in New York began weighing and measuring the bones of the *T. bataar* specimen in preparation for customs registration and shipment back

to Mongolia, which is expected shortly after it officially becomes Mongolian property on 6 May. The legal status of the fossil had been in limbo since Mongolia disputed its sale for more than US\$1 million at a New York auction house last May. The 7.3-metre-long specimen is one of the most complete tyrannosaurid specimens ever found.

day," said NASA launch commentator Kyle Herring. See go.nature.com/b6oeoz for more.

Lawsuit settlement

A cancer researcher has settled a lawsuit against the US Department of Health and Human Services (HHS) after successfully appealing against a finding of research misconduct. Philippe Bois, a former postdoctoral researcher at St Jude Children's Research Hospital in Memphis, Tennessee, denied that he committed research misconduct in the settlement. But he did not dispute that he inadvertently fabricated data in two research papers. The

Office of Research Integrity, part of the HHS, announced the settlement on 18 April. The case is the first time that a decision by a judge working for the HHS has been overturned.

Primate carriers

Vietnam Airlines said on 19 April that it will no longer transport primates used in research experiments, effective from 1 May. The airline has been under pressure from animal-rights groups. It was one of the last major carriers to transport primates for research: only Air France and Philippine Airlines say that they still do so. Air Canada, United Airlines and China

Eastern announced that they would stop shipments in December, January and March respectively.

Animal activism

Animal-rights activists occupied an animal facility at the University of Milan in Italy on 20 April. They demanded that all its 800 animals (mostly genetically modified mice) be transferred into their care. After 12 hours of negotiations, the activists agreed to leave with fewer than 100 animals, but mixed up some of the remaining animals and cage labels to disrupt experiments. Researchers say they have lost years of work. See go.nature.com/yxeciw for more.

PEOPLE

US energy secretary

In a vote on 18 April, the US Senate energy committee approved President Barack Obama's choice for energy secretary: Ernest Moniz, a physicist at the Massachusetts Institute of Technology in Cambridge. Moniz, who served as an undersecretary for energy under former president Bill Clinton, has backed the use of a mixture of conventional and renewable energy sources to meet demand (see *Nature* 494, 409–410; 2013). The full Senate is expected to confirm the nomination.

Nobel laureate dies

François Jacob, a Nobel-prizewinning French biologist, died on 19 April aged 92. With Jacques Monod and André Lwoff, Jacob shared the 1965 Nobel Prize in Physiology or Medicine for his work on gene expression and how it is controlled. While working at the Pasteur Institute in Paris, he identified regulatory proteins that bind to DNA, preventing its transcription into RNA and thus dampening the expression of cellular enzymes. Jacob explained how feedback from the cell's environment changes the activity of the regulatory proteins.



POLICY

African agriculture

African farmers must use sustainable and environmentally friendly technologies to reverse rising hunger levels across the continent, according to an 18 April report from the Montpellier Panel, a group of agriculture and development experts based in London. One recommended practice is pictured in Malawi: planting crops under 'fertilizer' trees, such as *Faidherbia albida*, which provide nutrients to the soil below. It says that sustainable intensification of African agriculture will produce higher yields and more nutritious foods while reducing reliance on fertilizers and pesticides, thus lowering greenhouse-gas emissions.

RESEARCH

Big bursts

Astronomers have spotted a new, long-lived and powerful type of γ -ray burst, a cosmic

explosion that spews out high-energy particles. The bursts can last for up to several hours at a time, rather than the seconds or minutes that scientists expected. They might emanate from the death throes of supergiant stars, astronomers proposed at a meeting in Nashville, Tennessee, on 16 April.

BUSINESS

Energy spending

Investment in renewable energy technologies still falls short of the level needed to clean up the global energy system and stabilize the climate, says a report from the International Energy Agency in Paris. In 2012, global markets in solar photovoltaic technology and wind energy grew by 42% and 19%, respectively, says the 17 April report. But the continued growth in energy produced by coal-fired power stations is offsetting progress, it says.

Venture declines

US venture-capital investments shrank 12% to US\$5.9 billion in the first quarter of 2013, with the life sciences and clean technology particularly affected, according to a report by accountancy firm PricewaterhouseCoopers in London and the National Venture Capital Association in Arlington, Virginia. The report,

COMING UP

24–25 APRIL

On World Malaria Day (25 April), scientists review research advances at the Johns Hopkins Malaria Research Institute in Baltimore, Maryland. go.nature.com/wfnw2

27–30 APRIL

Flu pandemics, the resurgence of measles and antimicrobial resistance are all discussed at the European Society of Clinical Microbiology and Infectious Diseases meeting in Berlin. go.nature.com/jyfhwf

released on 19 April, found that investment in biotechnology and medical devices fell by 28% and investment in clean technology declined by 35% relative to the previous quarter. First-time deals for start-ups in the life sciences dropped by 52% to \$98 million — the lowest level since 1996.

Salmon farming

The Haida Salmon Restoration Corporation (HSRC), a salmon breeding and biotechnology company on the Queen Charlotte Islands in Canada, is disputing the legality of a search of its offices by the government agency Environment Canada last month. The agency said that the corporation had dumped iron compounds off the west coast of Canada illegally. The HSRC says that the iron was intended to fertilize phytoplankton, boosting ocean productivity and salmon populations. On 17 April, the corporation filed a court brief arguing that Canadian anti-dumping regulations do not apply to "ocean pasture replenishment and restoration".

TREND WATCH

Prices for allowances to emit a tonne of carbon dioxide on Europe's carbon-trading market are likely to remain low until 2020, after the European Parliament rejected a plan on 16 April to withhold the release of some emissions allowances, which have flooded the market since the recession. This means that the market is unlikely to spur investment in low-carbon energy, one of the scheme's key goals when it was launched in 2005. See go.nature.com/czdx9k for more.

CARBON-MARKET COLLAPSE

Politicians voted against reviving prices in Europe's carbon-trading market — letting it slump to an all-time low of €2.7 (US\$3.5).



➔ **NATURE.COM**

For daily news updates see:
www.nature.com/news

NEWS IN FOCUS

ASTRONOMY A tug of war over naming rights for exoplanets **p.407**

ENVIRONMENT Europe nears decision on controversial pesticides **p.408**

GEOPHYSICS A two-billion-year journey through deep Earth **p.410**

PROFILE Seeking the truth about firearms and violence **p.412**



RICARDO AZOURY/REDUX/EYEVINE



The ways in which rising carbon dioxide levels will affect the Amazon rainforest are still highly uncertain.

CLIMATE CHANGE

Experiment aims to steep rainforest in carbon dioxide

Sensor-studded plots in the Amazon forest will measure the fertilizing effect of the gas.

BY JEFF TOLLEFSON

One of the wild cards in climate change is the fate of the Amazon rainforest. Will it shrivel as the region dries in a warming climate? Or will it grow even faster as the added carbon dioxide in the atmosphere spurs photosynthesis and allows plants to use water more efficiently? A dying rainforest could release gigatonnes of carbon into the atmosphere, accelerating warming; a CO₂-fertilized forest could have the opposite effect, sucking up carbon and putting the brakes on climate change.

Climate modellers trying to build carbon fertilization into their forecasts have had precious few data to go on. “The number one question is, how will tropical forests react if we put more CO₂ into the atmosphere?” says Carlos Nobre, a climate scientist who heads research programmes at the Brazilian Ministry of Science, Technology and Innovation in Brasilia. “We don’t know.”

Now an international team of scientists is developing an ambitious experiment in the central Amazon that could study

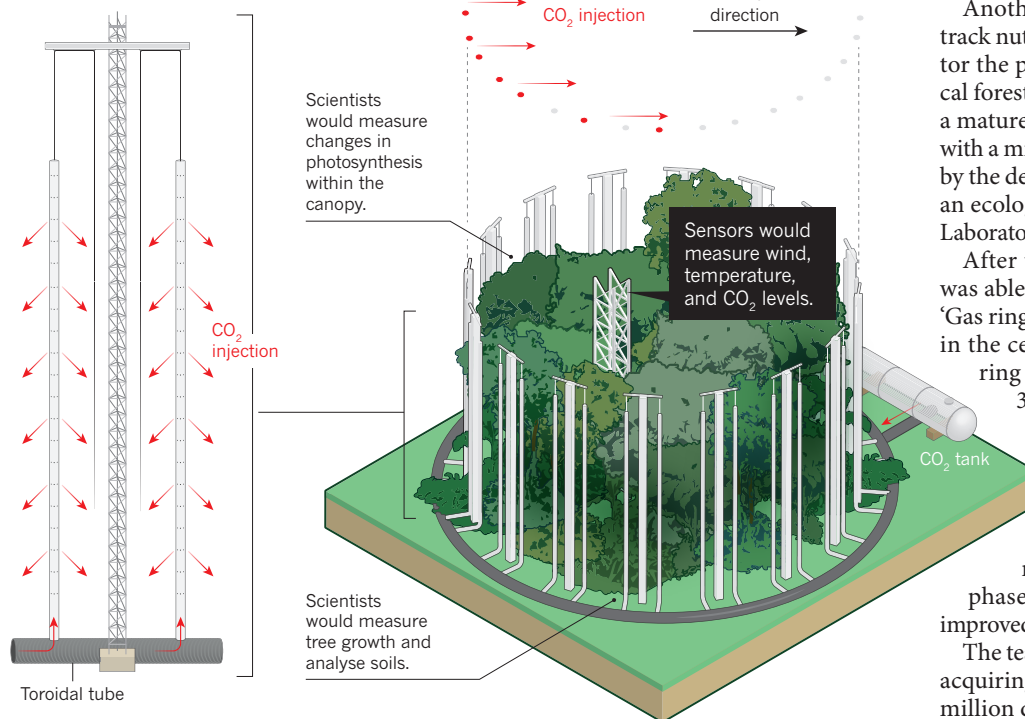
the effect in the real world. Hosted in Washington DC by the Inter-American Development Bank (IADB), a group of some 30 scientists met this month to flesh out the details of a project that would bathe a patch of rainforest in extra CO₂ and, over the course of a decade or more, measure how the plants respond. The experiment, the first of its kind in the tropics, would be modelled on free-air CO₂ enrichment (FACE) experiments conducted over the past couple of decades in the young and biologically simpler temperate forests of the Northern Hemisphere.

The experiment’s results could foretell the ►

NATURE.COM
For more on
Amazon ecology,
see:
go.nature.com/xfgmba

GAS RING

Scientists are planning an experiment in the Amazon rainforest that would measure how elevated carbon dioxide levels enhance plant growth.



► future of the Amazon. In 2000, a team at the UK Met Office's Hadley Centre in Bracknell proposed that drought caused by global warming could devastate the rainforest¹, although other climate modellers disagreed. Since then, the Hadley team has lowered its estimates of the likelihood of drying and the resulting forest dieback². But the Hadley Centre's simulations, like all climate models, assume a substantial CO₂-fertilization effect in the tropics.

In an atmosphere of elevated CO₂, not only do plants grow faster, but also their stomata (tiny openings on their leaves) do not need to open as widely or for as long. This means that less water escapes through transpiration, which makes plants better able to withstand heat and drought. The net result is that, at least in climate models, the extent of CO₂ fertilization largely determines the Amazon's resilience to global warming.

Because of the sheer volume of carbon cycling through the tropics, the fertilization effect has a massive impact on the amount of carbon that forests take up globally — and on how much remains in the atmosphere. Using the Hadley Centre climate model, UK modellers showed last year³ that atmospheric CO₂ levels in 2100 depended largely on the magnitude of the fertilization effect, and could vary from 669 to 1,130 parts per million (CO₂ levels today stand at 395 parts per million). That range corresponds to a 2.4°C rise in global temperatures. Richard Betts, a member of the Hadley team, says the paper showed that the effect of the CO₂-fertilization feedback was potentially

much larger than had been thought. "This is why we have been really keen for people to go out and study the Amazon forest," says Betts. "Our model indicates CO₂ enrichment, and we need to know how realistic it is."

The experiments in temperate forests — rings of towers that inject CO₂ into circular plots — showed an initial fertilization effect, although the long-term response varied depending on the availability of nutrients in the soil, such as nitrogen⁴. In theory, the fertilization effect should be stronger in the tropics, where warmer temperatures work in concert with higher CO₂ levels to increase the rate of photosynthesis (but plants shut down altogether if the temperature gets too high). Nitrogen is also more plentiful in the tropics, although other nutrients, such as phosphorus, could be limiting factors.

The idea of conducting a FACE experiment in the tropics has been around for years, but proposals have tended to fizzle out amid concerns about the feasibility of working in a mature tropical forest. First among them is the forest's diversity: how could an experiment be large enough to be representative of a forest that has thousands of species of canopy trees and a cascade of plants beneath? On this point, the scientists meeting in Washington simply threw up their hands. "At the end of the day, no experiment is representative of the totality of the biome," says Evan DeLucia, an ecologist

at the University of Illinois in Urbana-Champaign and one of the principal investigators in a FACE experiment on young pines in South Carolina.

Another challenge is developing tools to track nutrient cycles in the soil and to monitor the peculiar growth dynamics of tropical forests. For instance, most of the trees in a mature tropical forest are hardly growing, with a minority quickly filling in gaps created by the death of old trees, says Jeff Chambers, an ecologist at Lawrence Berkeley National Laboratory in Berkeley, California.

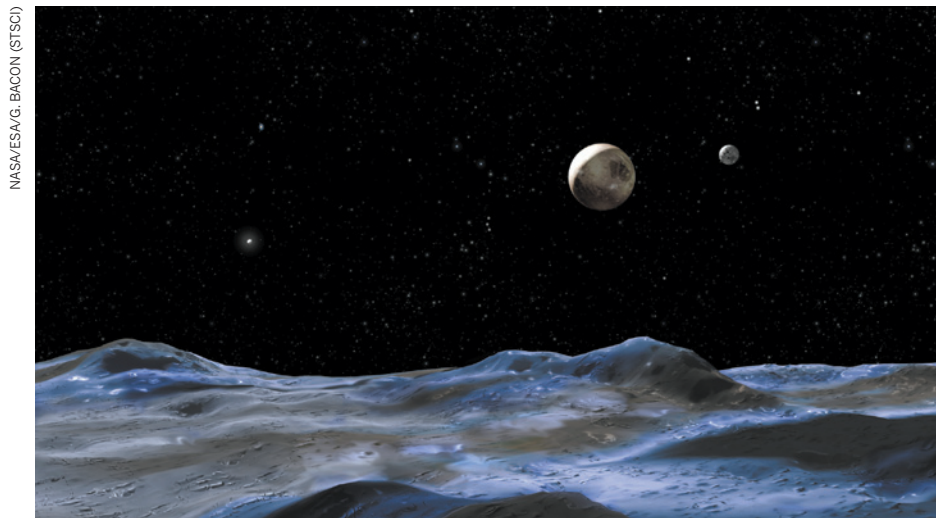
After two days of discussion, the group was able to converge on a basic design (see 'Gas ring'). A pilot project, north of Manaus in the central Amazon, would consist of a ring of about 16 towers circling an area 30 metres in diameter. Sensors would monitor background CO₂ levels and winds, and CO₂ would be injected from the towers as needed, to boost levels within the circle by 200 parts per million. Extra rings would be added in subsequent phases to allow replication and provide improved statistics.

The team is exploring different options for acquiring the CO₂, which could cost several million dollars annually for the full experiment. Options include buying the gas from a local beer and soft-drinks factory, and producing it independently, along with methane, from a local landfill. The team must also decide whether to build a pipeline for the CO₂ or to maintain a road for the large trucks that would deliver the gas. So daunting are the challenges that the team plans to ask the engineering arm of the Brazilian military for help.

The pilot project would cost about US\$10 million for the first few years, and scientists are looking to the IADB and the Brazilian Ministry of Science, Technology and Innovation for seed money. Nobre has encouraged the team to apply for a large grant from the Amazon Fund, a pot of money that Brazil uses to combat deforestation and promote sustainable development. The tentative goal is to begin fieldwork next year and to bring the pilot facility online in 2015.

Some scientists wonder whether the project will provide the answers they need at a price that politicians are willing to pay. But so far, its planners are finding themselves in the enviable position of being pushed to think big by potential funders. "Let's do it right," said Jerry Melillo, a senior scientist at the Marine Biological Laboratory in Woods Hole, Massachusetts, who proposed a FACE experiment in Brazil more than a decade ago. "We only get one chance." ■

1. Cox, P. M. et al. *Nature* **408**, 184–187 (2000).
2. Cox, P. M. et al. *Nature* **494**, 341–344 (2013).
3. Booth, B. B. B. et al. *Environ. Res. Lett.* **7**, 024002 (2012).
4. Norby, R. J., Warren, J. M., Iversen, C. M., Medlyn, B. E. & McMurtrie, R. E. *Proc. Natl Acad. Sci. USA* **107**, 19368–19373 (2010).



A view of Pluto imagined from one of its moons. Names are being considered for the fourth and fifth moons.

ASTRONOMY

Moon and planet names spark battle

Company clashes with International Astronomical Union over popular labels for exoplanets.

BY ALEXANDRA WITZE

Over breakfast one day in 1930, Falconer Madan, a librarian at the University of Oxford, UK, read a newspaper report about a newly discovered planet to his 11-year-old granddaughter. Little Venetia Burney piped up: why not name the new world Pluto? Madan passed the tip to a friend, a well-connected astronomer at the university. Within two months, what was then considered the ninth planet got its name.

Eight decades later, the public is still fascinated with naming other worlds, but the process of doing so can be contentious. In the coming weeks, the International Astronomical Union (IAU) in Paris, the organization in charge of nomenclature, is expected to decide on the names of the two most recently discovered

moons of Pluto, now classed as a dwarf planet.

The IAU is considering two names submitted by the discovery team: Vulcan and Cerberus, the top vote-winners in a popular contest run by the SETI Institute in Mountain View, California. *Star Trek* actor William Shatner suggested Vulcan, the runaway winner with more than 174,000 votes.

Even as people chuckle over the naming of Pluto's moons, tension is rising in another realm of celestial nomenclature: extrasolar planets. Nearly 1,000 exoplanet discoveries have been confirmed. With thousands more on the horizon, some scientists are saying that it is time for selected exoplanets to receive popular and easy-to-remember names, in addition to their technical codes, which do not exactly trip off the tongue. One example is HD 209458 b, a Jupiter-sized gas giant discovered in 1999 — the

first exoplanet to be found by watching for a dimming of its parent star's light.

In February, Uwingu, a space-education company based in Boulder, Colorado, launched a public contest, asking for contributions to what it calls "a baby book of names" for astronomers to draw on. Suggesting a name cost US\$4.99; voting was 99 cents. In March, Uwingu focused the contest to solicit names for the planet around the nearby star α Centauri B. Uwingu co-founder Alan Stern, a planetary scientist at the Southwest Research Institute in Boulder who helped to discover Pluto's newest moons, hopes that people will refer to the exoplanet by whichever name wins, regardless of whether the IAU endorses it.

Proceeds from Uwingu's competition, after expenses, are to be donated to space exploration and education projects, says Stern. By the time it ended on 22 April, the contest had received 1,242 nominations and 6,178 votes, for total earnings of around \$10,000. The winning name was Albertus Alauda, a Latinized version of the name of the nominator's late grandfather. But on 12 April, the IAU issued a press release titled "Can One Buy the Right to Name a Planet?". Without mentioning Uwingu by name, it re-asserted the IAU's role as the official astronomical namer. Uwingu countered with a release pointing out that many informal names for astronomical objects remain in use despite not getting IAU approval. For example, some astronomers have used Osiris as an informal name for HD 209458 b. And the IAU has sanctioned informal names for 17,766 of the Solar System's 360,190 catalogued asteroids.

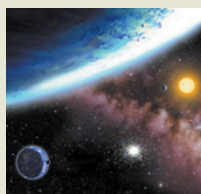
Alain Lecavelier des Etangs, an astronomer at the Paris Institute of Astrophysics and chairman of the IAU exoplanet-naming commission, says that the group expects to make a decision on whether to adopt popular names within the next six months.

The IAU's decision on the moons of Pluto is expected much sooner. However, the names Vulcan and Cerberus may not pass muster, because the IAU tries to avoid duplication. Vulcan is already the name of a hypothetical mini-planet once mooted to exist between Mercury and the Sun, and Cerberus is a 1.2-kilometre-wide asteroid. If the IAU rejects the proposed names, it will be up to Mark Showalter at the SETI Institute and his team, who discovered the moons in 2011 and 2012, to suggest alternatives. ■



**MORE
ONLINE**

TOP STORY

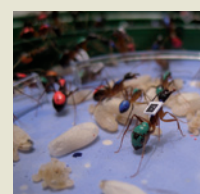


Two planets orbiting in life-friendly zone may be 'water worlds' go.nature.com/fukxor

MORE NEWS

- Hurricane could have triggered aftershocks go.nature.com/fctf1t
- Royal family might have evolved to cope with inbreeding go.nature.com/jgzvld
- Radioactive bacteria attack pancreatic-cancer metastases go.nature.com/r2s3sp

VIDEO



Whole-colony tracking shows that ants switch tasks as they age go.nature.com/jncoun

Europe debates risk to bees

Proposed pesticide ban gathers scientific support as some experts call for more field studies.

BY DANIEL CRESSEY

Across the globe, hives of honeybees are dying off in a phenomenon known as colony collapse disorder. Among the proposed culprits are pesticides called neonicotinoids, which are supposed to be less harmful to beneficial insects and mammals than the previous generation of chemicals.

Debate over neonicotinoids has become fierce. Conservation groups and politicians in the United Kingdom and Europe have called for a ban on their use, but agricultural organizations have said that farmers will face hardship if that happens. Next Monday, European governments will take a crucial vote on whether to severely restrict or ban three neonicotinoids.

Scientists, meanwhile, are vigorously debating whether the studies on neonicotinoids and the health of honeybees and bumblebees, mostly conducted in laboratory settings, accurately reflect what is happening to bees in the field.

Neonicotinoids, which poison insects by binding to receptors in their nervous systems, have been in use since the late 1990s. They are applied to crop seeds such as maize (corn) and soya beans, and permeate the plants, protecting them from insect pests. But a growing body of research suggests that sublethal exposure to the pesticides in nectar and pollen may be harming bees too — by disrupting their ability to gather pollen, return to their hives and reproduce¹⁻⁶ (see “The buzz over bee health”).

In January, the European Food Safety

Authority in Parma, Italy, Europe’s food-chain risk-assessment body, concluded that three commonly used neonicotinoids — clothianidin, imidacloprid and thiamethoxam — should not be used where they might end up in crops that attract bees, such as oilseed rape and maize. The European Commission then proposed a two-year ban on the use of these chemicals in such crops. That proposal failed to gain sufficient support last month in a vote by European Union member states, but on 29 April, ministers will vote again.

Some scientists say that there is insufficient evidence to implicate these compounds. Ecotoxicologist James Cresswell, who studies pollination at the University of Exeter, UK, says that “one can still equivocate over the evidence” because many of the lab studies that have shown harm may have fed bees unrealistically high doses of neonicotinoids. The problem, he adds, is that data are lacking on what doses bees actually encounter in the field. “Everyone is focused on hazard,” he says. “We know there is hazard there. But risk is a product of hazard and exposure.”

However, David Goulson, a bee researcher at the University of Sussex, UK, thinks that most of the major studies have used realistic doses. “I couldn’t say I am certain these impacts really occur in the field, but it seems to me very likely that they do,” he says.

Even if neonicotinoids are not directly responsible for colony collapse disorder, they could play a part by making bees more

susceptible to the parasitic mite *Varroa destructor* and the parasitic fungus *Nosema apis*, both prime suspects, adds Christian Krupke, an entomologist at Purdue University in West Lafayette, Indiana. He says that, on the basis of current evidence, neonicotinoid use should be restricted immediately as a precaution.

One of the few studies to be conducted in the field served only to stoke the controversy after its release in March⁶. Conducted by an agency within the UK Department for Environment, Food and Rural Affairs (DEFRA), it exposed 20 bumblebee colonies at three sites to crops grown from untreated, clothianidin-treated or imidacloprid-treated seeds. It found “no clear consistent relationships” between pesticide levels and harm to the insects.

DEFRA also reviewed the body of evidence on neonicotinoids and concluded that, although there might be “rare effects of neonicotinoids on bees in the field”, these do not occur under normal circumstances.

Experts lined up to criticize the field study. Neuroscientist Christopher Connolly of the University of Dundee, UK, who has studied the effect of neonicotinoids in bee brains, says that the control colonies themselves were contaminated with the pesticides, and that thiamethoxam was detected in two of the three bee groups tested, even though it was not used in the experiment. Goulson agrees, saying of the study: “In many ways, it was appalling.” No one from DEFRA was available to talk to *Nature*.

Goulson and others say that intensive environmental monitoring of neonicotinoids and long-term field studies of their effects are sorely needed. He points to a 2012 study⁷ that found neonicotinoids in dandelions growing near treated crops, suggesting that the pesticides can spread from their intended target. “This debate has focused very heavily on bees. Perhaps we’re missing a slightly bigger picture,” he says. “For 20 years we’ve been using neonicotinoids without really assessing what impact they might be having in the wider environment.” ■

1. Henry, M. *et al. Science* **336**, 348–350 (2012).
2. Whitehorn, P. R., O’Connor, S., Wackers, F. L. & Goulson, D. *Science* **336**, 351–352 (2012).
3. Gill, R. J., Ramos-Rodriguez, O. & Raine, N. E. *Nature* **491**, 105–108 (2012).
4. Williamson, S. M. & Wright, G. A. J. *Exp. Biol.* <http://doi.org/k2z> (2013).
5. Palmer, M. J. *et al. Nature Commun.* **4**, 1634 (2013).
6. Thompson, H. *et al. Effects of Neonicotinoid Seed Treatments on Bumble Bee Colonies Under Field Conditions* (Food and Environment Research Agency, 2013).
7. Krupke, C. H., Hunt, G. J., Eitzer, B. D., Andino, G. & Given, K. *PLoS ONE* **7**, e29268 (2012).

INSECTICIDE EFFECTS

The buzz over bee health

The past year has seen a raft of papers about the effects of neonicotinoid pesticides on bees. Scientists are debating their real-world significance.

20 April 2012: Honeybees in French fields exposed to thiamethoxam show impaired homing back to hives¹. And bumblebee colonies exposed to “field-realistic levels” of imidacloprid in labs show a decreased growth rate and an 85% reduction in new queen production, compared with controls².

21 October 2012: “Field-level exposure” of bumblebees to imidacloprid and a non-neonicotinoid insecticide impairs foraging, increases worker-bee mortality and reduces colony success³.

7 February 2013:

“Prolonged exposure” to imidacloprid and another insecticide impairs learning and memory in honeybees⁴.

27 March 2013:

Lab study shows that imidacloprid, clothianidin and an organophosphate pesticide block firing of honeybee brain cells, especially when combined⁵.

March 2013: “No clear consistent relationships” seen between neonicotinoid levels and colony mass or production of new queens by bumblebee hives⁶. **D.C.**



MARK BOWLER/NATUREPL.COM



Methane being burnt off at sea after a team in Japan extracted the gas from frozen offshore deposits.

ENERGY

Japanese test coaxes fire from ice

First attempt to extract methane from frozen hydrates far beneath the ocean shows promise.

BY DAVID CYRANOSKI

Methane flowing from beneath the sea floor has buoyed Japan's hopes for securing its own plentiful energy source. A pilot project 80 kilometres off the country's shores produced tens of thousands of cubic metres of gas — and reams of useful data — before a clogged pump brought the project to an abrupt end last month.

Reservoirs of methane hydrates — icy deposits in which methane molecules are trapped in a lattice of water — are thought to hold more energy than all other fossil fuels combined. The problem is extracting the methane economically from the deposits, which lie beneath Arctic permafrost and seafloor sediments. But some scientists and policy-makers in energy-poor, coast-rich Japan hope that the reservoirs will become a crucial part of the country's energy profile.

Engineers have had some limited success in extracting methane from underneath Canadian tundra. But tapping the richer marine deposits presents a host of challenges, among them the

fact that whereas oil and natural gas exist in deep reservoirs, methane hydrates are found in the first few hundred metres of the sea bottom where sediments are loose, making wells unstable and putting them at risk of clogging by sand.

The test, run by the Tokyo-based state oil company Japan Oil, Gas and Metals National Corporation (JOGMEC), took place in waters 1 kilometre deep, where the research drilling ship *Chikyu* had bored through 270 metres of sediment to reach a 60-metre-thick methane hydrate reservoir. On 12 March, a pump reduced the pressure in the deposit, unlocking the gas from its icy cage. Gas started flowing up from the sea floor to a platform on the ship, where it produced a roaring flame. "Being Japanese, you might have thought we would have yelled 'banzai' or something," says project director Koji Yamamoto. But he says that he was too busy staring at displays of crucial data showing the pressure at the bottom of the well and the flow rate and composition of the incoming gas.

The big question, and the one on which Japanese energy hopes depend, is whether the engineers can sustain the flow. They did — for a while. The methane flowed smoothly for six days, with the flow rate increasing as the pressure dropped, generating an average of 20,000 cubic metres a day — more than Yamamoto expected and ten times more than was produced by a well dug in Canadian permafrost in 2008 using the same depressurization method.

It is "a remarkable breakthrough," says Scott Dallimore, a geoscientist at the Geological Survey of Canada in Sidney, British Columbia, who worked on the Canadian project with JOGMEC but was not involved in the Japanese offshore test. "The engineering challenge — to successfully undertake the test in a marine setting — was not insignificant. The flow rates are also very encouraging," he says.

Ray Boswell, technology manager for the methane hydrates programme at the US Department of Energy's National Energy Technology Laboratory in Morgantown, West Virginia, says that the test demonstrates that "what we have learned in the Arctic can be transferred to the marine environment, where the most significant resources are". From his experience of extracting methane from hydrates in Alaska, the team would have had to overcome significant obstacles, he says: the loose, shifting sediment, unpredictable weather and the fact that the methane cools its surroundings as it dissociates from the ice slush, potentially creating new hydrates that could slow production or clog up the well.

Yamamoto says that his team took care to avoid such problems. To stop the formation of icy hydrates, the researchers carefully lowered the pressure in the reservoir, aiming to cap it at 3 megapascals (MPa) by the end of the two-week test to keep the methane in gas form. But on the sixth day, with the pressure down to 4.5 MPa, the pump clogged up with sand and the test had to stop. "It was a disappointment," says Yamamoto. The team had used two sifting devices to try to prevent such a clog.

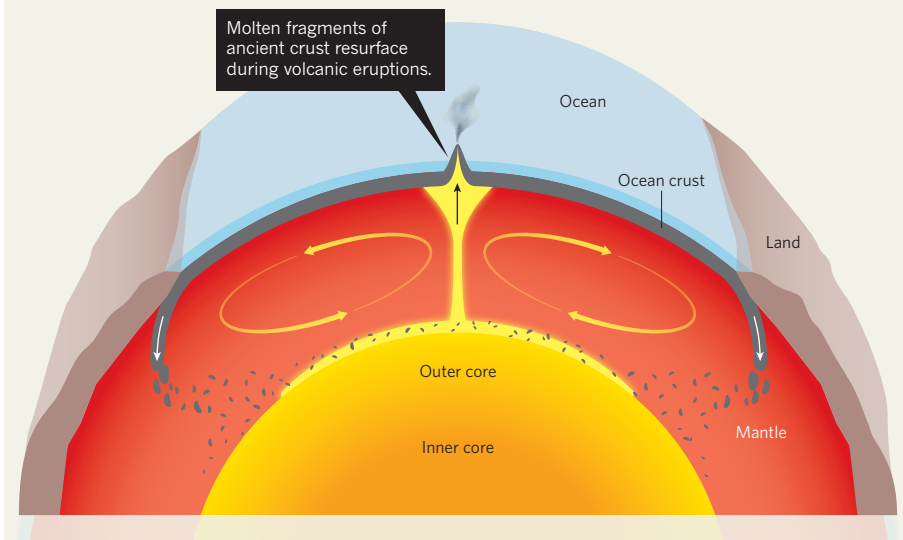
Yamamoto is confident that this and other obstacles can be overcome to create a steady supply of methane, but adds that improved extraction technologies and higher flow rates will be key to making the enterprise economically feasible. "We are 10 or 20 years behind shale, before they came up with fracking," he says. Others are not sure it is worth it. Canada and the United States have drastically cut their methane hydrate efforts, largely because they have plentiful gas from shale. Projects in China, India and South Korea, however, remain active.

The team will now examine temperature, seismic and other data to learn how far the dissociation of hydrates spread and thus how much methane they might expect to extract from one well. Yamamoto plans to spend a year preparing the next test, which he hopes will run for a further 12 months and will use more sophisticated monitoring. ■

➔ **NATURE.COM**
For more on
extracting methane
hydrates, see:
go.nature.com/mobk76

RETURN FROM THE CORE

Analyses of 20-million-year-old volcanic rocks from a remote island (centre) suggest they contain remnants of Earth's crust that sank or were forced deep into the mantle more than 2.45 billion years ago.



GEOPHYSICS

Ancient crust rises from the deep

Remnants of surface rocks take long tour of planet's interior.

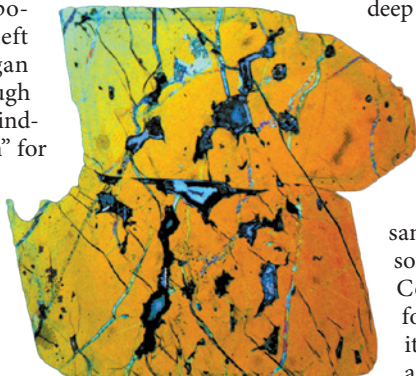
BY SID PERKINS

Earth recycles — but it takes its time. Chemical remnants of the rigid surface plates that plunge deep into the planet's interior at subduction zones can eventually resurface on distant volcanic islands. But the process may take more than two billion years, a study published in this issue¹ suggests.

By analysing volcanic rock that erupted millions of years ago on an island in the South Pacific, the researchers found clues about when components of the rock first left Earth's surface and began their long journey through its interior. The authors' findings are "a smoking gun" for deep, slow tectonic recycling, says Steven Shirey, a geochemist at the Carnegie Institution for Science in Washington DC. "It's hard to conclude that they're not right."

Studies of volcanic rock have revealed that the

chemical and isotopic composition of Earth's mantle — the layer of molten rock beneath the crust — varies considerably from place to place, says Rita Cabral, a geochemist at Boston University in Massachusetts and a co-author of the paper. Some have proposed that those variations arose because chunks of crust that once resided at Earth's surface have tainted parts of the mantle^{2,3}. But researchers have had to rely on computer models to estimate how fast the recycling takes place — and firm evidence that material is recycled through the planet's deep interior has been lacking.



Olivine crystals hold chemical clues to their origins.

Cabral and her colleagues now have compelling evidence that such tectonic recycling really happens, and of how long it takes¹. The team analysed rock samples from Mangaia, the southernmost of Polynesia's Cook Islands. The rocks, formed by volcanic activity about 20 million years ago, have been worn by weathering. But sulphide minerals locked away inside

weather-resistant crystals of olivine, which formed at a depth of a few kilometres before spewing from the volcano, still retain their pre-eruption composition, says Cabral.

That composition is telling. For one thing, Cabral notes, the proportion of the isotope sulphur-33 is substantially lower than that typically found in Earth's crust. Although biological processes can generate such an anomaly, they would simultaneously generate abnormally high concentrations of sulphur-34 — which are not present in the Mangaia samples.

The most likely source of the sulphur-33-depleted rocks, the team says, is mantle material that includes remnants of crust that sank or were pushed below Earth's surface at least 2.45 billion years ago, before photosynthetic organisms filled the atmosphere with oxygen. When oxygen was low, sunlight-driven reactions would naturally have created sulphides containing lower-than-normal proportions of sulphur-33; later, the ozone layer resulting from the surge of oxygen would have stifled those reactions.

At some point, Cabral contends, material from the core-mantle boundary upwelled in a 'hotspot' — a large-scale version of the buoyancy-driven bubbling seen in the lava lamps that were popular during the 1970s (see 'Return from the core'). The upwelling swept the sulphur-33-depleted material back to the surface.

In addition to providing insight into the pace of tectonic recycling, the findings reveal how little violent mixing occurs deep within Earth, says Cabral. The purported piece of ancient crust containing the sulphur-33-depleted minerals "had to have stayed relatively intact in the mantle for all that time", she notes, implying that the deep mantle may be a graveyard of ancient tectonic slabs.

Shirey sees a broader implication: that modern-style plate tectonics were in motion at least 2.45 billion years ago. That's a conclusion that some researchers resist, arguing that the young planet still had too much internal heat for surface plates to have been subducted into the mantle, as they are today.

"This is exciting, and there's no doubt there's recycling of ancient material," says Robert Stern, a geoscientist at the University of Texas at Dallas. But he suggests that sulphur-33-depleted material might have formed not at the surface but on the underside of a section of continental crust and then "dripped" down into the mantle, a process that some seismic studies suggest may be happening in certain regions today.

The case for ancient plate tectonics is far from closed, Stern says. "When plate tectonics began and what was happening before that time are still open questions." ■

1. Cabral, R. A. *et al.* *Nature* **496**, 490–493 (2013).
2. Hofmann, A. W. & White, W. M. *Earth Planet. Sci. Lett.* **57**, 421–436 (1982).
3. White, W. M. & Hofmann, A. W. *Nature* **296**, 821–825 (1982).

J. M. D. DAY

ETHICS

Guidance issued for US Internet research

Institutional review boards may need to take a closer look at some types of online research.

BY ERIKA CHECK HAYDEN

Andrew Gordon studies the way that people narrate events in their lives. The computer scientist, who is based at the University of Southern California in Los Angeles, has a seemingly inexhaustible source of raw data for his experiments: blogs. And, although the authors of these blogs often obscure their identities, Gordon says that it is relatively easy to figure out who they are, by using information from photographs that they post or by looking up the registrant of the blog's domain name.

Can Gordon use information from the blog posts freely? As the Internet has become an ever-more essential research tool, scientists and institutional review boards (IRBs) facing such questions have been frustrated by the muddiness of existing regulations.

Now, an advisory committee to the US Department of Health and Human Services (DHHS), which governs human-subjects research, has endorsed a 20-point set of recommendations that could help. But some scientists worry that the recommendations might place more areas of Internet research under the purview of IRBs, which have been attacked by their critics as capricious, overly cautious groups that add time, complexity and costs to studies (see A. Halavais *Nature* **480**, 174–175; 2011).

Although the DHHS secretary has not officially endorsed the recommendations, administrators say it is already being used. "People are going to use this whether it gets blessed officially beyond this committee or not, because it is so urgently needed," says Susan Rose, the University of Southern California's executive director for the protection of research subjects.

In some instances, the new recommendations could help IRBs to be less cautious. For example, the document clarifies when investigators must verify the identities of research participants, an issue that has bedeviled IRBs. The guidelines suggest that, in a low-risk study such as an online survey, a check box, confirming that respondents are accurately representing themselves, could be sufficient. But for studies that could seriously impact person's well-being — for instance, a clinical trial — researchers might need to obtain proof of age and identity, and require participants

to pass a quiz to show that they understand the research.

The guidelines also suggest that, in general, information on the Internet should be considered public, and thus not subject to IRB review — even if people falsely assume that the information is anonymous. Yet the guidelines complicate the issue by suggesting that IRB review might be needed if there are doubts about the work's 'beneficence' — the idea that all research should be conducted with the welfare of its subjects in mind. For instance, a clinical-trial manager should not recruit patients from an online disease support group, says Elizabeth Buchanan, chair of the Center for Applied Ethics at the University of Wisconsin–Stout. "There are places where individuals may have a reasonable expectation of privacy based on the context of the site," says Buchanan, a co-author of the guidelines.

Gordon's work might, at first glance, seem to be exempt from IRB review, because he analyses public blogs accessible to anyone.

"There are places where individuals may have a reasonable expectation of privacy."

But a closer read of the new recommendations suggests that Gordon does need to get IRB review. The bloggers he studies don't realize that their identities are readily available, and they could be harmed if some of the details they discuss were to be publicly linked to them by researchers. "There's a gap between the expectation and the reality of what can be done with technology, so it really complicates the issue of what is identifiable private information," Gordon says.

Some worry that the guidelines may put entirely new areas of research under the purview of IRBs. The document briefly suggests that researchers' Twitter streams and blogs might be subject to IRB review if they are used for patient recruitment. That bothers Don Dizon, an oncologist at Massachusetts General Hospital in Boston who has served on IRBs for nine years. "All of a sudden IRBs are not only protecting human research subjects, but they're also policing their own investigators, which is not an efficient way to use time or money," he says. ■

THE GUN FIGHTER



BY MEREDITH WADMAN

THERE ARE ALMOST AS MANY
FIREARMS IN THE UNITED STATES
AS THERE ARE CITIZENS.
GAREN WINTEMUTE IS ONE OF
FEW PEOPLE STUDYING THE
CONSEQUENCES.

With his crisp blue suit and wire-framed spectacles, Garen Wintemute hardly looked frightening as he stepped to the podium last month to address a conference on paediatric emergency medicine in San Francisco, California. But his presence there made the organizers nervous.

Wintemute, an emergency-department doctor, is better known as the director of the Violence Prevention Research Program at the University of California (UC), Davis. As such, he has published dozens of papers on the effects of guns in the United States, where widespread gun ownership and loose laws make it easy for criminals and potentially violent people to obtain firearms. Wintemute has pushed the bounds of research, going undercover into gun shows with a hidden camera to document how people often sidestep the law when purchasing weapons. He has also worked with California lawmakers on crafting gun policy and helped to drive a group of gun-making companies out of business.

All this made Wintemute a potentially risky speaker for the conference funder, a branch of the US Department of Health and Human Services, which is barred by law from funding any activities that advocate or promote gun control. The meeting organizers had told Wintemute to stick to facts and avoid any mention of policies. But with the nation still reeling from the murder of 20 children and 6 educators, who were shot in their school in Newtown, Connecticut, in December, the conference organizers were not sure what Wintemute would say.

He stuck to the facts, but also managed to make clear how he feels about the funding prohibition, which has effectively killed off most research on gun violence. "We don't have a labour force," Wintemute told the assembled doctors.

That has led to a striking imbalance in US medical research. Firearms accounted for more than 31,000 deaths in the United States in 2011 (see 'Gun deaths'). But fewer than 20 academics in the country study gun violence, and most of them are economists, criminologists or sociologists. Wintemute is one of just a few public-health experts devoted to this research, which he has funded through a mixture of grants and nearly US\$1 million of personal money.

His undercover gun-show tactics have led him into situations where he feared for his safety, and they have also raised protests from some gun-rights advocates, who charge that Wintemute is more a biased campaigner than a researcher.

But even a few of his ideological opponents praise Wintemute's work. "Garen is one of the very best in terms of his research skills," says David Kopel, the research director at the Independence Institute in Denver, Colorado, a think tank that supports gun-owners' rights.

And Wintemute, who is 61, makes no apologies for his passion or his methods. "I believe

just as strongly as I can articulate in the value of free inquiry," he says, "especially when the stakes are so high — when so many people are dying through no fault of their own; when so much of the country simply turns its back on this problem."

AIMING TRUE

Wintemute grew up in a home in Long Beach, California, where his father, a decorated veteran of the Second World War, kept a Japanese officer's sabre and infantry rifle, a Winchester carbine and a Marlin .22 calibre rifle in a bedroom cupboard. Wintemute learned to shoot, and begged to go hunting. That chance came when he was around 12, and his father asked him to help clear out sparrows from the rafters of his company's warehouse.

Wintemute's aim was good, he recalls. "But I held those birds and looked at the finality of it all and felt them turn cold in my hands and decided this was not for me."

As an undergraduate at Yale University in New Haven, Connecticut, Wintemute flirted with oceanography and neuroscience, but eventually decided that he wanted to be a physician. After completing medical school and a residency in family practice, both at UC Davis, Wintemute went to work in 1981 as medical coordinator at the Nong Samet Refugee Camp, just inside Cambodia's border with Thailand. The camp was in an area that had only recently been liberated from the Khmer Rouge dictator Pol Pot, and Wintemute took care of gunshot wounds on a daily basis. Even more common were shrapnel injuries from land mines. There was no electricity, and amputations were done under local anaesthetic.

"I never once met an intact family," Wintemute recalls. "Everybody had lost somebody. There came a point where I said: 'I need to pick up a rifle. I can't be on the sidelines.'"

But instead of grabbing a gun, Wintemute decided to pursue 'big-picture' international health. He left Cambodia and enrolled in a one-year master's programme in public health at Johns Hopkins University in Baltimore, Maryland. One of his first courses was taught by a former trial lawyer named Stephen Teret, who is now director of the Center for Law and the Public's Health at Johns Hopkins.

Teret remembers the day in September 1982 when the students of that class introduced themselves and Wintemute stunned him with his charisma and eloquence. "I said to myself: 'I'm going to get to know this guy,'" recalls Teret, and the two of them soon became friends and collaborators.

On a cold winter day several months later, some close friends of Teret's dropped their 21-month-old son off at the house of his caregiver. Around noon, the caregiver

laid him down for a nap and left the room, whereupon her four-year-old son took his father's loaded handgun from a nearby drawer, pointed it at the sleeping infant and shot him through the head.

Within weeks, Teret switched his main research focus from motor-vehicle injuries to gun injuries, an area in which public-health research was all but non-existent. Wintemute began assisting him, and their first project was a law-review article laying out a legal strategy for suing gun-makers who fail to use available safety technologies to prevent accidental gun deaths¹.

Wintemute returned to UC Davis, with the goal of focusing on gun injuries. In Cambodia and then in the Sacramento emergency department, Wintemute learned the hard lesson that, as a doctor, he had little chance of saving many people with gunshot wounds;

most of those who died did so before they even reached the hospital. He realized that if he wanted to reduce deaths from firearms, he needed to prevent shootings in the first place.

One day, he set himself a question as he left for a run in the foothills east of Sacramento. Looking to make an impact, he wondered: "What subset of firearm injuries can people simply not turn away from?" By the time he got back, he had decided to focus on the kind of shooting that had shattered the lives of Teret's friends.

In June 1987, Wintemute published a paper called 'When children shoot children: 88 unintended deaths in California'². He reported that in 36% of these cases, the shooters didn't think that the gun was loaded or was real, or they were too young to tell the difference. Forty per cent of the children's fatal injuries were self-inflicted, including separate incidents in which a 5-year-old boy and a 2-year-old boy, using .38-calibre revolvers — one found under a pillow, the other in his parents' bedroom — each shot himself in the head.

To illustrate one facet of the problem, Wintemute borrowed several of the guns used in the shootings from the Sacramento medical examiner. He then bought toy lookalikes, mounted the paired guns on a piece of plywood and, when the paper was published, called a press conference. Few of the reporters who attended could tell the toy guns from the real ones. His work and other events that year focused scrutiny on toy guns, and in December, toy retailers began to pull realistic-looking toy guns from their shelves. The next year, California banned their sale and manufacture.

Wintemute was increasingly convinced that gun manufacturing was a pressure point that could be turned to advantage, by tying the industry to the public-health consequences

"SO MUCH OF THE COUNTRY
SIMPLY TURNS ITS BACK
ON THIS PROBLEM."

of its products. He was contemplating how to do that when the *Wall Street Journal* published an article about a group of companies in and around Los Angeles, California, owned by one extended family that made small-calibre, inexpensive handguns known as Saturday Night Specials. Poorly made and lacking some safety features, the guns were disproportionately used in crime, particularly by juveniles.

The article contained a trove of details about the family that ran the companies, and Wintemute decided to follow that trail. The result was *Ring of Fire*, a book published in 1994 that described the enterprise and impact of the six companies, which in 1992 produced 34% of the handguns made in the country.

Ring of Fire painted such a stark portrait of the problematic guns that “it became the focus of the rallying cry for local legislative action”, says Sayre Weaver, a lawyer who represented West Hollywood, the first of several Los Angeles communities to ban the sale of the Saturday Night Specials. In 1999, the California legislature followed by making it illegal to manufacture and sell the handguns. Within several years, 5 of the 6 companies were out of business.

BATTLE TO SURVIVE

Although his book had a big impact, Wintemute's research soon hit a snag. With grant support from the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, Wintemute had been conducting a retrospective cohort study looking at whether handgun buyers with prior misdemeanour convictions are more likely than those without a criminal history to be charged with new crimes, particularly those involving firearms and violence. (Many states allow purchases by criminals who have been convicted of misdemeanours, such as assault.)

But as he was digging into the study, his source of funding came under attack from the National Rifle Association (NRA), a powerful lobbying group based in Fairfax, Virginia, that supports gun ownership. NRA leaders were upset with the CDC for funding work by another researcher who had found that people with a gun in their home were 2.7 times more likely than those without to be murdered³, and 4.8 times more likely to commit suicide⁴.

In 1996, the NRA persuaded congressman Jay Dickey (Republican, Arkansas) to insert language into a budget bill to prohibit the CDC from advocating or promoting gun control. (That ban has been renewed every year since then.) Dickey's amendment also stripped \$2.6 million from the agency's 1997 funding — the exact amount that the CDC had spent on firearm research the previous year.

In 1996, Wintemute had received \$292,000 from the CDC for the misdemeanour study, but after the change, the agency provided just \$50,000 to close down the programme.

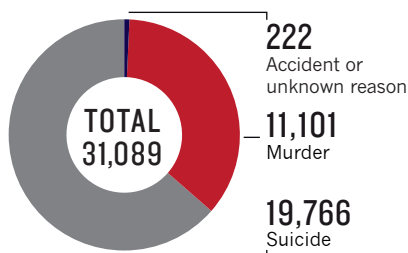
The research restrictions were extended

in 2012 to encompass all of the CDC's parent agency, the Department of Health and Human Services. And they have had a measurable effect. According to an analysis of Elsevier's Scopus database by the group Mayors Against Illegal Guns, the proportion of all publications dealing with US firearms and their impacts declined by 60% between 1996 and 2010.

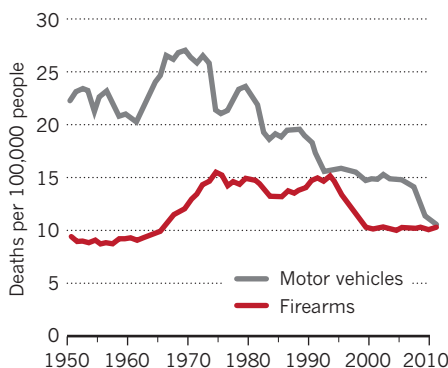
US researchers still produce more papers per capita on the topic than do investigators

GUN DEATHS

Firearms accounted for 1.2% of US deaths in 2011, with suicides being the largest fraction.



Unlike deaths from car accidents, the rate of gun fatalities has flattened out. Research restrictions have hampered efforts to explain the gun trend.



from other countries. But the subject may not be as high on other countries' research agendas because gun ownership is so much lower in most developed nations (see “Top gun”). The United Kingdom, for example, banned private possession of handguns in 1998 after a gunman shot and killed 16 children and their teacher in a school in Dunblane, Scotland¹.

Wintemute was rare in staying devoted to gun research after the restrictions were imposed. He turned to the California Wellness Foundation, a large private charity based in Woodland Hills that focuses on health care and health education, and the foundation provided the funds to complete his study. Wintemute followed up nearly 6,000 authorized handgun purchasers, most of them for 15 years. He found that men who had had two or more convictions for misdemeanour violence were 15 times as likely as those with no criminal history to be charged with the most violent crimes⁵.

Today, Wintemute runs the four-person Violence Prevention Research Program at UC

Davis, on about \$300,000 a year, none of which comes from the federal government. Of this, \$50,000 is from the California Wellness Foundation. Until last year, Wintemute also received substantial funding from both the California and US departments of justice. Since 2005, he has donated \$945,000 from his own savings and stock sales to the programme.

In July, the university announced that it would endow two professor slots to support Wintemute's programme, each of which comes with \$75,000 a year. Wintemute has assumed one and is looking to fill the other one, a position in violence epidemiology.

The hiring comes at a time of renewed activity in the field. After the December school shooting, President Barack Obama ordered the CDC to resume research into the causes of gun violence and the ways to prevent it; his 2014 budget request, released on 10 April, asks Congress to provide \$10 million for the research. This week in Washington DC, Wintemute spoke to an Institute of Medicine panel that has been formed to advise the CDC on which research questions are most pressing.

INSIDE OUT

As Wintemute delved into gun research in the 1980s, he decided to immerse himself in the gun culture. He joined the NRA and the rifle and pistol club in Davis, where he practised shooting at an indoor range. In 1999, he started to visit gun shows, good opportunities to observe firearm purchases. “Gun shows are sort of like zoos,” he says. “You can easily see a wide range of behaviours.”

At his first show in Milwaukee, Wisconsin, the signs used to advertise guns caught his attention. One licensed retailer displayed a Mossberg Model 500 shotgun with a pistol grip next to a poster that read “Great for Urban Hunting”. Another sign, beside a Savage rifle, read: “Great for Getto [sic] Cruisers”.

Wintemute says that he was astonished by the blatant promotion of guns as murder weapons. “It was clearly a story that had to be told — bearing witness is part of the job — but I wanted to figure out a way to tell the story quantitatively, scientifically.”

It took several years of trial and error at shows before he was confident enough of his methods to begin collecting data. He cut off his waist-length ponytail so he would not stand out in the crowds, bought a small camera and placed it in a bag of Panda liquorice with a lens-sized hole cut in the side. A pen and notepad would attract too much notice, so he set up his office voicemail so that he could call it from his mobile phone and record long messages. He later added a video camera disguised to look like a button on his shirt.

Several times, Wintemute was accused of taking unauthorized photos, and his phone was temporarily confiscated by security personnel, who examined it and found no pictures. After one such episode, he says, a

colleague overheard a group of men planning to attack Wintemute outside the show, but Wintemute successfully avoided them.

Altogether, he attended 78 gun shows in 19 states, strolling the aisles while apparently deep in a phone conversation. A paper on the findings showed, among other things, that the restrictive policies regulating gun shows in California resulted in fewer illegal 'straw' purchases — in which someone buys a gun on behalf of a person legally barred from doing so — than in other states⁶.

By 2008, Wintemute was contending with being outed: David Codrea, the author of a blog called WarOnGuns, had posted Wintemute's photo online with the note: "WARNING! IF YOU SEE THIS MAN, NOTIFY SECURITY IMMEDIATELY." The post identified Wintemute by name and called him an "anti-gun 'researcher'" who stalked gun shows with hidden cameras and recorders.

But by that point, Wintemute says, he had learned all he could and stopped going to shows.

CRITICAL APPROACH

Last month, on the day after Wintemute spoke to the emergency researchers in San Francisco, the NRA posted a critique slamming a study⁷ that reported that states with more firearm laws had lower rates of firearm fatalities.

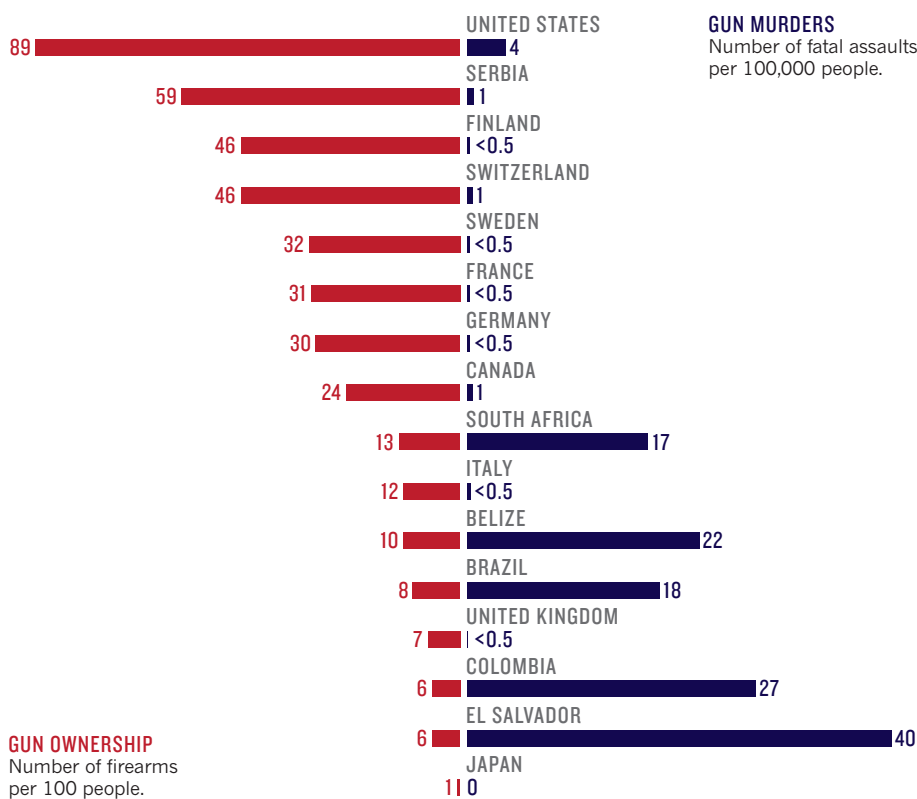
The NRA quoted from an unlikely source to attack the paper: Wintemute, who had published a sharp rebuttal to the paper in the same journal⁸. Wintemute had argued that the association between more laws and fewer deaths disappeared when the authors accounted for firearm ownership in a state — meaning that it is impossible to say whether the restrictive gun laws save lives by inhibiting gun ownership or whether laws are simply easier to enact in states in which ownership rates are already low. The latter is a more plausible explanation, he wrote.

One of the paper's authors, Eric Flegler, an emergency physician at Boston Children's Hospital in Massachusetts, responds that "when you look at firearm-related homicides, even controlling for firearm ownership, firearm-related homicides do decrease in states with more gun laws".

This is not the first time that Wintemute has attacked papers he perceives to be weak, even if they point towards policies he would like to see adopted. And he goes no easier on policies that he views as ineffective, even ones that seek to limit firearm ownership. He has, for instance, repeatedly criticized the assault-weapons ban enacted by Congress in 1994, in part because the ban was easily circumvented. Instead, he advocates three steps informed by research: requiring background checks for all US gun sales, forbidding alcohol abusers and those convicted of violent misdemeanours from buying guns and rewriting current federal restrictions on gun ownership to better capture people who are mentally ill and at risk of violence to

TOP GUN

The United States has the most firearms per capita and the greatest number of gun murders of any developed nation.



GUN OWNERSHIP

Number of firearms per 100 people.

themselves or others.

Wintemute's rigour has earned the respect of some ideological opponents, but others say that his work betrays anti-gun biases by, for instance, selectively citing the literature in a way that minimizes the value of firearms for self-defence.

"We have followed his research for many years. Pro-gun scholars have criticized it for just as long," says John Frazer, director of the Research and Information Division at the NRA's lobbying arm, the Institute for Legislative Action in Fairfax.

Wintemute's work at gun shows has also triggered complaints. Kopel, the Independence Institute's researcher, says that Wintemute's hidden-camera tactics were "sleazy". "I have a higher opinion of him as a guy who looks at the data and analyses them in a serious way," Kopel says.

Now, Wintemute is focusing on a new project. He is designing a randomized trial to study roughly 20,000 people who purchased guns legally in California but have since lost the right to own firearms because they committed a violent crime, were served with a domestic-violence restraining order or were judged mentally ill and potentially violent. Unlike in other states, authorities in California have begun to take guns away from those people. Wintemute is hoping to test the effectiveness of the policy by comparing re-offence rates among those whose guns are seized quickly versus those who keep them for longer.

The money for his own work, at least in the short term, will probably have to come from California or from private sources. Wintemute is not optimistic that funds for CDC firearm research will be forthcoming from Congress in the short term.

Whether or not the federal money materializes, Wintemute will continue the work he began 30 years ago. For him, it is part of his mission as a physician to relieve suffering. "Everything that was true of firearm violence in the early 1980s is still true today," he says. "There is a fundamental injustice in violence. People don't ask for it; it comes to them." ■

Meredith Wadman is a reporter for *Nature* in Washington DC.

1. Teret, S. P. & Wintemute, G. J. *Hamline Law Rev.* **6**, 341–350 (1983).
2. Wintemute, G. J., Teret, S. P., Kraus, J. F., Wright, M. A. & Bradfield, G. J. *Am. Med. Assoc.* **257**, 3107–3109 (1987).
3. Kellermann, A. L. *et al.* *N. Engl. J. Med.* **329**, 1084–1091 (1993).
4. Kellermann, A. L. *et al.* *N. Engl. J. Med.* **327**, 467–472 (1992).
5. Wintemute, G. J., Drake, C. M., Beaumont, J. J., Wright, M. A. & Parham, C. A. *J. Am. Med. Assoc.* **280**, 2083–2087 (1998).
6. Wintemute, G. J. *Inj. Prev.* **13**, 150–155 (2007).
7. Flegler, E. W., Lee, L. K., Monuteaux, M. C., Hemenway, D. & Mannix, R. *J. Am. Med. Assoc. Intern. Med.* <http://dx.doi.org/10.1001/jamainternmed.2013.1286> (2013).
8. Wintemute, G. J. *J. Am. Med. Assoc. Intern. Med.* <http://dx.doi.org/10.1001/jamainternmed.2013.1292> (2013).

Schizophrenia

Bipolar disorder

On the spectrum

BY DAVID ADAM

Pyromania

Social phobia

Research suggests that mental illnesses lie along a spectrum — but the field's latest diagnostic manual still splits them apart.

David Kupfer is a modern-day heretic. A psychiatrist at the University of Pittsburgh in Pennsylvania, Kupfer, has spent the past six years directing the revision of a book commonly referred to as the bible of the psychiatric field. The work will reach a climax next month when the American Psychiatric Association (APA) unveils the fifth incarnation of the book, called the *Diagnostic and Statistical Manual of Mental Disorders* (DSM), which provides checklists of symptoms that psychiatrists around the world use to diagnose their patients. The DSM is so influential that just about the only suggestion of Kupfer's that did not meet with howls of protest during the revision process was to change its name from *DSM-V* to *DSM-5*.

Although the title and wording of the manual are now settled, the debate that overshadowed the revision is not. The stark fact is that no one has yet agreed on how best to define and

diagnose mental illnesses. *DSM-5*, like the two preceding editions, will place disorders in discrete categories such as major-depressive disorder, bipolar disorder, schizophrenia and obsessive-compulsive disorder (OCD). These categories, which have guided psychiatry since the early 1980s, are based largely on decades-old theory and subjective symptoms.

The problem is that biologists have been unable to find any genetic or neuroscientific evidence to support the breakdown of complex mental disorders into separate categories. Many psychiatrists, meanwhile, already think outside the category boxes, because they see so many patients whose symptoms do not fit neatly into them. Kupfer and others wanted the latest DSM to move away from the category approach and towards one called 'dimensionality', in which

➔ **NATURE.COM**

For more on challenges in mental-health research, see: go.nature.com/6xgksp

mental illnesses overlap. According to this view, the disorders are the product of shared risk factors that lead to abnormalities in intersecting drives such as motivation and reward anticipation, which can be measured (hence 'dimension') and used to place people on one of several spectra. But the attempt to introduce this approach foundered, as other psychiatrists and psychologists protested that it was premature.

Research could yet come to the rescue. In 2010, the US National Institute of Mental Health (NIMH) in Bethesda, Maryland, launched an initiative, called the Research Domain Criteria project, that aims to improve understanding of dimensional variables and the brain circuits involved in mental disorders. Clinical psychologist Bruce Cuthbert, who heads the project, says that it is an attempt to go "back to the drawing board" on mental illness. In place of categories, he says, "we do have to start thinking instead about how these disorders are dysregulation in normal processes".

But that will be too late for the *DSM*. Kupfer says that he now sees how hard it is to change clinical doctrine. "The plane is in the air and we have had to make the changes while it is still flying."

MANUAL EVOLUTION

The Catholic Church changes its pope more often than the APA publishes a new *DSM*. The first and second editions, published in 1952 and 1968, reflected Sigmund Freud's idea of psychodynamics: that mental illness is the product of conflict between internal drives. For example, *DSM-I* listed anxiety as "produced by a threat from within the personality". Symptoms were largely irrelevant to diagnosis.

Things got more empirical around 1980. Shocked by the discovery that patients with identical symptoms were receiving different diagnoses and treatments, an influential group of US psychiatrists threw out Freud and imported another role model from central Europe: psychiatrist Emil Kraepelin. Kraepelin famously said that the conditions now known as schizophrenia and bipolar disorder were separate syndromes, with unique sets of symptoms and presumably unique causes. *DSM-III*, published in 1980, turned this thinking into what is now called the category approach, with solid walls between conditions. When the existing version, *DSM-IV*, came out in 1994, it simply added and subtracted a few categories.

Since then, an entire generation of troubled individuals has trooped into psychiatric clinics and left with a diagnosis of a *DSM*-approved condition, including anxiety disorder, eating disorders and personality disorders. Most of those conditions will appear in the pages of *DSM-5*, the contents of which are officially under wraps until the APA annual meeting — which starts in San Francisco, California, on 18 May — but have been an open secret since the APA published a draft on its website last year and invited comment.

But even as walls between conditions were being cemented in the profession's manual, they were breaking down in the clinic. As psychiatrists well know, most patients turn up with a mix of symptoms and so are frequently diagnosed with several disorders, or co-morbidities. About one-fifth of people who fulfil criteria for one *DSM-IV* disorder meet the criteria for at least two more.

These are patients "who have not read the textbook", says Steve Hyman, who directs the

**"We need to
give researchers
permission to
think outside these
traditional silos."**

Stanley Center for Psychiatric Research, part of the Broad Institute in Cambridge, Massachusetts. As their symptoms wax and wane over time, they receive different diagnoses, which can be upsetting and give false hope. "The problem is that the *DSM* has been launched into under-researched waters, and this has been accepted in an unquestioning way," he says.

Psychiatrists see so many people with co-morbidities that they have even created new categories to account for some of them. The classic Kraepelian theoretical division between schizophrenia and bipolar disorder, for example, has long been bridged by a pragmatic hybrid called schizoaffective disorder, which describes those with symptoms of both disorders and was recognized in *DSM-IV*.

Basic research has offered little clarification. Despite decades of work, the genetic, metabolic and cellular signatures of almost all mental syndromes remain largely a mystery. Ironically, the ingrained category approach is actually inhibiting the scientific research that could refine diagnoses, in part because funding agencies have often favoured studies that fit the standard diagnostic groups. "Until a few years ago we simply would not have been able to get a grant to study psychoses," says Nick Craddock, who works at the Medical Research Council Centre for Neuropsychiatric Genetics and Genomics at Cardiff University, UK. "Researchers studied bipolar disorder or they studied schizophrenia. It was unthinkable to study them together."

"We need to give researchers permission to think outside these traditional silos," says Hyman. "We need to get them to re-analyse these conditions from the bottom up."

In the past few years, some researchers have taken up the challenge — and the findings from genetics and brain-imaging studies support the idea that the *DSM* disorders overlap. Studies with functional magnetic resonance imaging

show that people with anxiety disorders and those with mood disorders share a hyperactive response of the brain's amygdala region to negative emotion and aversion¹. Similarly, those with schizophrenia and those with post-traumatic stress disorder both show unusual activity in the prefrontal cortex when asked to carry out tasks that require sustained attention¹.

And in the largest study yet undertaken to try to pinpoint the genetic roots of mental disorder, a group led by Jordan Smoller at the Massachusetts General Hospital in Boston screened genome information from more than 33,000 people with five major mental-health syndromes, looking for genetic sequences associated with their illness². At the end of February, the team reported that some genetic risk factors — specifically, four chromosomal sites — are associated with all five disorders: autism, attention deficit hyperactivity disorder, bipolar disorder, major depression and schizophrenia. "What we see in the genetics mirrors what we see in the clinic," Hyman says. "We are going to have to have a rethink."

RIVAL APPROACH

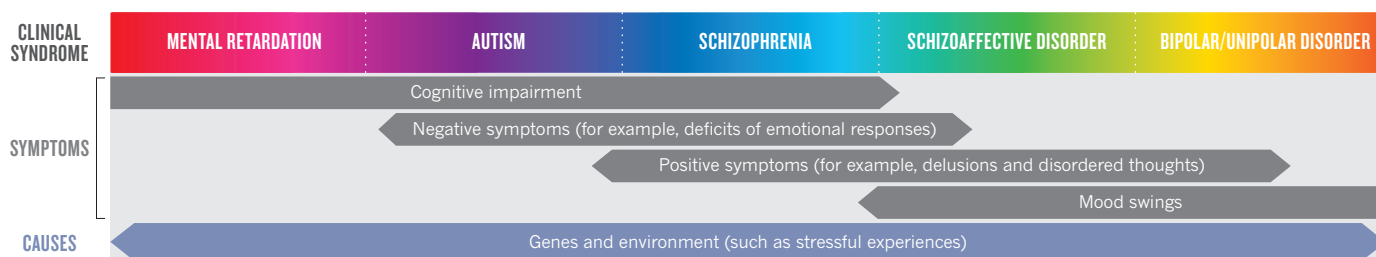
At the same time that research and clinical practice are helping to undermine the *DSM* categories, the rival dimensional approach is gaining support. Over the past decade, psychiatrists have proposed a number of such dimensions, but they are not used in practice — partly because they are not sanctioned by the *DSM*.

The frequent co-morbidity between schizophrenia and OCD, for instance, has led some to suggest a schizo-obsessive spectrum, with patients placed according to whether they attribute intrusive thoughts to an external or internal source. And in 2010, Craddock and his colleague Michael Owen proposed the most radical dimensional spectrum so far³, in which five classes of mental disorder are arranged on a single axis: mental retardation–autism–schizophrenia–schizoaffective disorder–bipolar disorder/unipolar mood disorder (see 'Added dimensions'). Psychiatrists would place people on the scale by assessing the severity of a series of traits that are affected in these conditions, such as cognitive impairment or mood disruption. It is a massively simplified approach, Craddock says, but it does seem to chime with the symptoms that patients report. More people show the signs of both mental retardation and autism, for example, than of both mental retardation and depression.

When Kupfer and his *DSM-5* task force began work in 2007, they were bullish that they would be able to make the switch to dimensional psychiatry. "I thought that if we did not use younger, more-basic science to push as hard as we could, then we would find it very difficult to move beyond the present state," Kupfer recalls. The task force organized a series of conferences to discuss how the approach could be introduced. One radical and particularly controversial proposal was to scrap half of the existing ten

Added dimensions

In the dimensional approach to psychiatry, mental-health conditions lie on a spectrum (example shown here) that has partly overlapping causes and symptoms.



conditions relating to personality disorder and introduce a series of cross-cutting dimensions to measure patients against, such as degree of compulsivity.

But this and other proposals met with stinging criticism. The scales proposed were not based on strong evidence, critics said, and psychiatrists had no experience of how to use them to diagnose patients. What is more, the personality-disorder dimensions flopped when they were tested on patients in field trials of the draft *DSM* criteria between 2010 and 2012: too many psychiatrists who tried them reached different conclusions. “Introducing a botched dimensional system prematurely into *DSM-5* may have the negative effect of poisoning the well for their future acceptance by clinicians,” wrote Allen Frances, emeritus professor of psychiatry at Duke University in Durham, North Carolina, in an article in the *British Journal of Psychiatry*⁴. Frances had served as head of the *DSM-IV* task force and was one of the strongest critics of proposals to introduce dimensionality to *DSM-5*.

The proposal was also unpopular with patient groups and charities, many of which have fought long and hard to make various distinct mental-health disorders into visible brands. They did not want to see schizophrenia or bipolar disorder labelled as something different. Speaking privately, some psychologists also mutter about the influence of drug companies and their relationship with psychiatrists. Both stand to profit from the existing *DSM* categories because health-insurance schemes in the United States pay for treatments based on them. They have little incentive to see categories dissolve.

CHANGE OF TACK

In the middle of 2011, the *DSM-5* task force admitted defeat. In an article in the *American Journal of Psychiatry*⁵, Kupfer and Darrel Regier, vice-chair of the *DSM-5* task force and the APA’s research director, conceded that they had been too optimistic. “We anticipated that these emerging diagnostic and treatment advances would impact the diagnosis and classification of mental disorders faster than what has actually occurred.” The controversial personality-disorder dimensions were voted down by the APA’s board of trustees at the final planning meeting in December 2012.

The APA claims that the final version of

DSM-5 is a significant advance on the previous edition and that it uses a combination of category and dimensional diagnoses. The previously separate categories of substance abuse and substance dependence are merged into the new diagnosis of substance-use disorder. Asperger’s syndrome is bundled together with a handful of related conditions into the new category called autism-spectrum disorder; and OCD, compulsive hair-pulling and other similar disorders are grouped together in an obsessive-compulsive and related disorders category. These last two changes, Regier says, should help research scientists who want to look at links between conditions. “That probably won’t make much difference to treatment but it should facilitate research into common vulnerabilities,” he says.

The Research Domain Criteria project is the biggest of these research efforts. Last year, the NIMH approved seven studies, worth a combined US\$5 million, for inclusion in the project — and, Cuthbert says, the initiative “will represent an increasing proportion of the NIMH’s translational-research portfolio in years to come”. The goal is to find new dimensional variables and assess their clinical value, information that could feed into a future *DSM*.

One of the NIMH-funded projects, led by Jerzy Bodurka at the Laureate Institute for Brain Research in Tulsa, Oklahoma, is examining anhedonia, the inability to take pleasure from activities such as exercise, sex or socializing. It is found in many mental illnesses, including depression and schizophrenia.

Bodurka’s group is studying the idea that dysfunctional brain circuits trigger the release of inflammatory cytokines and that these drive anhedonia by suppressing motivation and pleasure. The scientists plan to probe these links using analyses of gene expression and brain scans. In theory, if this or other mechanisms of anhedonia could be identified, patients could be tested for them and treated, whether they have a *DSM* diagnosis or not.

One of the big challenges, Cuthbert says, is to get the drug regulators on board with the idea that the *DSM* categories are not the only way to prove the efficacy of a medicine. Early talks about the principle have been positive, he says. And there are precedents: “Pain is not a disorder and yet the FDA gives licences for anti-pain drugs,” Cuthbert says.

Going back to the drawing board makes

sense for the scientists, but where does it leave *DSM-5*? On the question of dimensionality, most outsiders see it as largely the same as *DSM-IV*. Kupfer and Regier say that much of the work on dimensionality that did not make the final cut is included in the section of the manual intended to provoke further discussion and research. *DSM-5* is intended to be a “living document” that can be updated online much more frequently than in the past, Kupfer adds. That’s the reason for the suffix switch from V to 5; what comes out next month is really *DSM-5.0*. Once the evidence base strengthens, he says, perhaps as a direct result of the NIMH project, dimensional approaches can be included in a *DSM-5.1* or *DSM-5.2*.

All involved agree on one thing. Their role model now is not Freud or Kraepelin, but the genetic revolution taking place in oncology. Here, researchers and physicians are starting to classify and treat cancers on the basis of a tumour’s detailed genetic profile rather than the part of the body in which it grows. Those in the psychiatric field say that genetics and brain imaging could do the same for diagnoses in mental health. It will take time, however, and an entire generation will probably have to receive flawed diagnoses before the science is developed enough to consign the category approach to clinical history.

“I hope I’ll be able to give a patient with possible bipolar a proper clinical assessment,” Craddock says. “I’ll do a blood test and look for genetic risks and send them into a brain scanner and ask them to think of something mildly unhappy to exercise their emotional system.” The results could be used to trace the underlying cause — such as a problematic chemical signal in the brain. “I’ll then be able to provide lifestyle advice and treatment.” He pauses. “Actually it won’t be me, because I will have retired by then.” ■ SEE EDITORIAL P. 397

David Adam is Nature’s Editorial and Columns editor.

1. Dichter, G. S., Damiano, C. A. & Allen, J. A. *J. Neurodev. Disord.* **4**, 19 (2012).
2. Cross-Disorder Group of the Psychiatric Genomics Consortium *Lancet* [http://dx.doi.org/10.1016/S0140-6736\(12\)62129-1](http://dx.doi.org/10.1016/S0140-6736(12)62129-1) (2013).
3. Craddock, N. & Owen, M. J. *Br. J. Psychiatry* **196**, 92–95 (2010).
4. Frances, A. *Br. J. Psychiatry* **195**, 391–392 (2009).
5. Kupfer, D. J. & Regier, D. A. *Am. J. Psychiatry* **168**, 672–674 (2011).

COMMENT

RECOVERY Lab devastated by Hurricane Sandy picking up the pieces six months on **p.421**

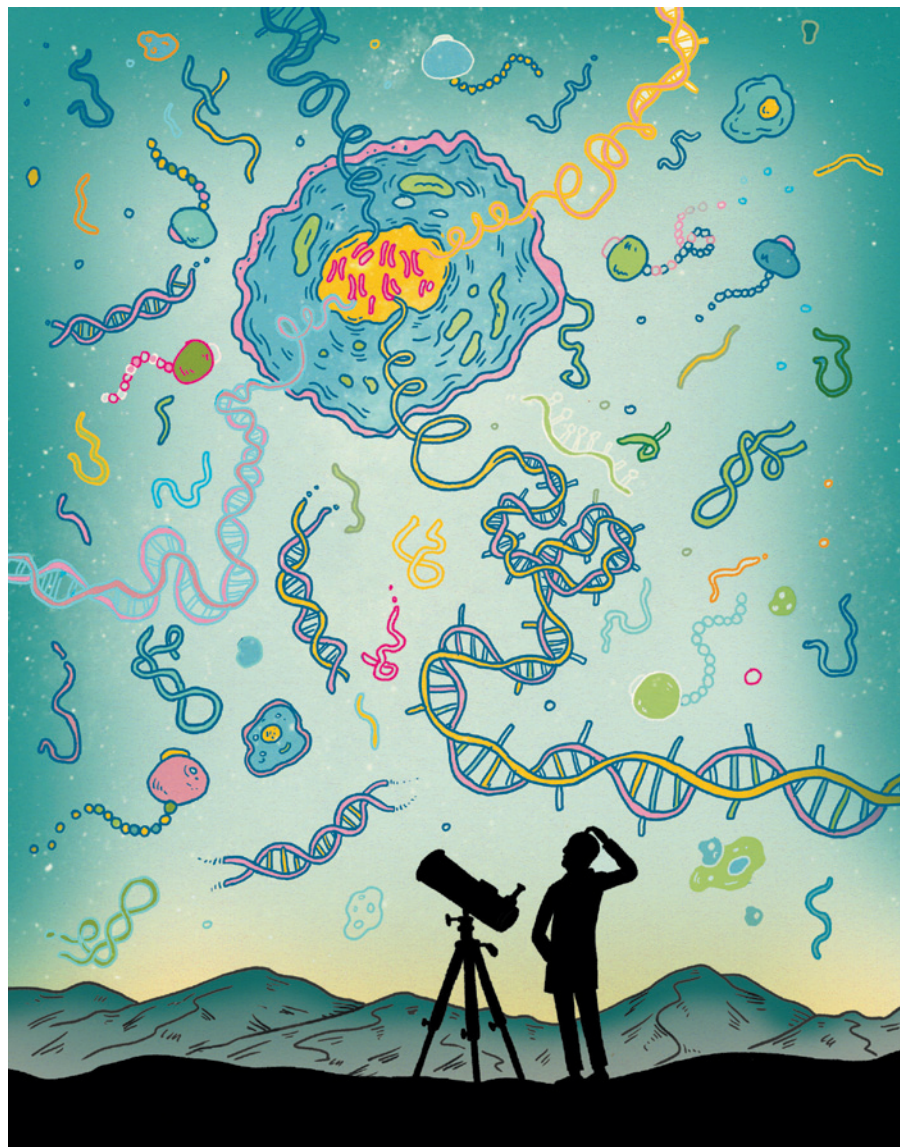
NEUROSCIENCE Spring Books special opens with Douglas Hofstadter's latest **p.424**

CULTURE Steve Jones' new book explores the science invoked by the Bible **p.432**



DNA Newfound letter sheds light on nominations behind double-helix Nobel **p.434**

ANDREW RAE



Celebrate the unknowns

On the 60th anniversary of the double helix, we should admit that we don't fully understand how evolution works at the molecular level, suggests **Philip Ball**.

This week's diamond jubilee of the discovery of DNA's molecular structure rightly celebrates how Francis Crick, James Watson and their collaborators launched the 'genomic age' by revealing how hereditary information is encoded in the double helix. Yet the conventional narrative — in which their 1953 *Nature* paper led inexorably to the Human Genome Project and the dawn of personalized medicine — is as misleading as the popular narrative of gene function itself, in which the DNA sequence is translated into proteins and ultimately into an organism's observable characteristics, or phenotype.

Sixty years on, the very definition of 'gene' is hotly debated. We do not know what most of our DNA does, nor how, or to what extent it governs traits. In other words, we do not fully understand how evolution works at the molecular level.

That sounds to me like an extraordinarily exciting state of affairs, comparable perhaps to the disruptive discovery in cosmology in 1998 that the expansion of the Universe is accelerating rather than decelerating, as astronomers had believed since the late 1920s. Yet, while specialists debate what the latest findings mean, the rhetoric of popular discussions of DNA, genomics and evolution remains largely unchanged, and the public continues to be fed assurances that DNA is as solipsistic a blueprint as ever.

The more complex picture now emerging raises difficult questions that this outsider knows he can barely discern. But I can tell that the usual tidy tale of how 'DNA makes RNA makes protein' is sanitized to the point of distortion. Instead of occasional, muted confessions from genomics boosters and popularizers of evolution that the story has turned out to be a little more complex, there should be a bolder admission — indeed a celebration — of the known unknowns.

DNA DISPUTE

A student referring to textbook discussions of genetics and evolution could be forgiven for thinking that the 'central dogma' devised by Crick and others in the 1960s — in which information flows in a linear, traceable fashion from DNA sequence to messenger RNA to protein, to manifest finally as phenotype — remains the solid foundation of the genomic revolution. In fact, it is beginning to look more like a casualty of it. ►

► Although it remains beyond serious doubt that Darwinian natural selection drives much, perhaps most, evolutionary change, it is often unclear at which phenotypic level selection operates, and particularly how it plays out at the molecular level.

Take the Encyclopedia of DNA Elements (ENCODE) project, a public research consortium launched by the US National Human Genome Research Institute in Bethesda, Maryland. Starting in 2003, ENCODE researchers set out to map which parts of human chromosomes are transcribed, how transcription is regulated and how the process is affected by the way the DNA is packaged in the cell nucleus. Last year, the group revealed¹ that there is much more to genome function than is encompassed in the roughly 1% of our DNA that contains some 20,000 protein-coding genes — challenging the old idea that much of the genome is junk. At least 80% of the genome is transcribed into RNA.

Some geneticists and evolutionary biologists say that all this extra transcription may simply be noise, irrelevant to function and evolution². But, drawing on the fact that regulatory roles have been pinned to some of the non-coding RNA transcripts discovered in pilot projects, the ENCODE team argues that at least some of this transcription could provide a reservoir of molecules with regulatory functions — in other words, a pool of potentially ‘useful’ variation. ENCODE researchers even propose, to the consternation of some, that the transcript should be considered the basic unit of inheritance, with ‘gene’ denoting not a piece of DNA but a higher-order concept pertaining to all the transcripts that contribute to a given phenotypic trait³.

According to evolutionary biologist Patrick Phillips at the University of Oregon in Eugene, projects such as ENCODE are showing scientists that they don’t really understand how genotypes map to phenotypes, or how exactly evolutionary forces shape any given genome.

COMPLEX CODE

The ENCODE findings join several other discoveries in unsettling old assumptions. For example, epigenetic molecular alterations to DNA, such as the addition of a methyl group, can affect the activity of genes without altering their nucleotide sequences. Many of these regulatory chemical markers are inherited, including some that govern susceptibility to diabetes and cardiovascular disease⁴. Genes can also be regulated by the spatial organization of the chromosomes, in turn affected by epigenetic markers. Although such effects have long been known, their prevalence may be much greater than previously thought⁵.

Another source of ambiguity in the genotype–phenotype relationship comes from the way in which many genes operate in

complex networks. For example, many differently structured gene networks might result in the same trait or phenotype⁶. Also, new phenotypes that are viable and potentially superior may be more likely to emerge through tweaks to regulatory networks than through more risky alterations to protein-coding sequences⁷. In a sense this is still natural selection pulling out the best from a bunch of random mutations, but not at the level of the DNA sequence itself.

One consequence of this complex genotype–phenotype relationship is that it may impose constraints on natural selection. If the same phenotypes can result from many similarly structured gene networks, it might take a long time for a ‘fitter’ phenotype to arise⁸. Alternatively, mutations may accumulate, free from selective ‘weeding’, thanks to the robustness of networks in maintaining a particular phenotype. Such hidden variation might be unmasked by some new environmental stress, enabling fresh adaptations to emerge⁹. These sorts of constraints and opportunities are poorly understood; evolutionary theory does not help biologists to predict what kinds of genetic network they should expect to see in any one context.

Researchers are also still not agreed on whether natural selection is the dominant driver of genetic change at the molecular level. Evolutionary geneticist Michael Lynch of Indiana University Bloomington has shown through modelling that random genetic drift can play a major part in the evolution of genomic features, for example the scattering of non-coding sections, called introns, through protein-coding sequences. He has also shown that rather than enhancing fitness, natural selection can generate a redundant accumulation of molecular ‘defences’, such as systems that detect folding problems in proteins¹⁰. At best, this is burdensome. At worst, it can be catastrophic.

In short, the current picture of how and where evolution operates, and how this shapes genomes, is something of a mess. That should not be a criticism, but rather a vote of confidence in the healthy, dynamic state of molecular and evolutionary biology.

A PROBLEM SHARED

Barely a whisper of this vibrant debate reaches the public. Take evolutionary biologist Richard Dawkins’ description in *Prospect* magazine last year of the gene as a replicator with “its own unique status as a unit of Darwinian selection”. It conjures up the decades-old picture of a little, autonomous stretch of DNA intent on getting itself copied, with no hint that selection operates at all levels

of the biological hierarchy, including at the supraorganismal level², or that the very idea of ‘gene’ has become problematic.

Why this apparent reluctance to acknowledge the complexity? One roadblock may be sentimentality. Biology is so complicated that it may be deeply painful for some to relinquish the promise of an elegant core mechanism. In cosmology, a single, shattering fact (the Universe’s accelerating expansion) cleanly rewrote the narrative. But in molecular evolution, old arguments, for instance about the importance of natural selection and random drift in driving genetic change, are now colliding with questions about non-coding RNA, epigenetics and genomic network theory. It is not yet clear which new story to tell.

Then there is the discomfort of all this uncertainty following the rhetoric surrounding the Human Genome Project, which seemed to promise, among other things, ‘the instructions to make a human’. It is one thing to revise our ideas about the cosmos, another to admit that we are not as close to understanding ourselves as we thought.

There may also be anxiety that admitting any uncertainty about the mechanisms of evolution will be exploited by those who seek to undermine it. Certainly, popular accounts of epigenetics and the ENCODE results have been much more coy about the evolutionary implications than the developmental ones. But we are grown-up enough to be told about the doubts, debates and discussions that are leaving the putative ‘age of the genome’ with more questions than answers. Tidying up the story bowdlerizes the science and creates straw men for its detractors. Simplistic portrayals of evolution encourage equally simplistic demolitions.

When the structure of DNA was first deduced, it seemed to supply the final part of a beautiful puzzle, the solution for which began with Charles Darwin and Gregor Mendel. The simplicity of that picture has proved too alluring. For the jubilee, we should do DNA a favour and lift some of the awesome responsibility for life’s complexity from its shoulders. ■

Philip Ball is a freelance science writer based in London.

e-mail: p.ball@btinternet.com

1. The ENCODE Project Consortium *Nature* **489**, 57–74 (2012).
2. Doolittle, W. F. *Proc. Natl Acad. Sci. USA* **110**, 5294–5300 (2013).
3. Djebali, S. et al. *Nature* **489**, 101–108 (2012).
4. Jablonka, E. & Raz, G. *Q. Rev. Biol.* **84**, 131–176 (2009).
5. Mattick, J. S. *Proc. Natl Acad. Sci. USA* **109**, 16400–16401 (2012).
6. Wagner, A. *Trends Genet.* **27**, 397–410 (2011).
7. Mattick, J. S. *FEBS Lett.* **585**, 1600–1616 (2011).
8. Wagner, A. *Trends Ecol. Evol.* **26**, 577–584 (2011).
9. Jarosz, D. F. & Lindquist, S. *Science* **330**, 1820–1824 (2010).
10. Lynch, M. *Proc. Natl Acad. Sci. USA* **109**, 18851–18856 (2012).



One of the many streets of Manhattan that flooded and lost power after the storm surge in New York in October 2012.

After the deluge

Gordon Fishell describes how he rebuilt his mouse research programme following the devastation wrought by Hurricane Sandy.

Six months ago, nearly all of my lab animals at the New York University (NYU) School of Medicine in Manhattan died. On 29 October 2012, Hurricane Sandy swept through the northeast coast of the United States. Salty water from the East River broke into the basement of my building, drowning 3,000 mice that carried 80 different traits I was studying. The mouse colony — which I used to study how neurons communicate with other cells — had been built up over 20 years. Many of my colleagues experienced similarly catastrophic losses.

Had I known exactly when and where the storm would pass, and just how bad it would be, I would have done more to prepare. We knew that a hurricane was coming, so we left the lab assuming that no one would be there for a few days. We put things away and checked that the emergency power was on. The animal-care people gave our mice extra water and food; we couldn't move the mice, because they had to stay in a germ-free environment to avoid infections, and there was nowhere large enough to put them.

On the day of the storm, all mass

transportation was closed, so I was forced to stay at home in Westchester. We knew early in the day that the water was rising — the park behind my house was flooded, so my son and I paddled around in our canoe. When I checked the weather report at 5 p.m., I saw that the storm had started tracking directly over the medical-school campus, and was going to hit in about two hours — at high tide. We were done for. It was obvious that our labs were in great danger, and there was nothing I could do.

The next day, I went stir-crazy at home — telephones and power were down, so there was no way of finding out how bad it was. My colleague Daniel Turnbull tried to drive us into Manhattan, but the bridges were closed. While in the car, I had limited phone reception and called Goichi Miyoshi, a postdoc who had made it into the lab. He told me that the generator had failed so the power was out, but that many crucial elements — cell lines, DNA constructs, primers and so on — were safe because he and other lab members had arrived at 7 a.m. and begun moving items to another building that had electricity.

“What about the mice?” I asked. “Are they okay?”

“They're all dead,” he said.

We turned the car around and went back home. I felt an awful sense of despair, for the suffering and loss of the animals, for the years of work lost and for the impact this would have on the people in my lab who had put their hearts and souls into their research. I mourned for 12 hours, then realized that I needed to work out how to move forward.

DAMAGE LIMITATION

When I finally reached the NYU medical-school campus, two days after the storm, it was organized chaos. There were trucks everywhere carrying dry ice and liquid nitrogen, and a loud buzz from the huge generators supplying emergency power to some of the buildings. Inside the medical centre where I work, the temperatures were exactly the same as outside — around 10°C. People were in shock but pulling together in a heartening way. Dafna Bar-Sagi, senior vice-president and vice-dean for science, who had barely slept in two days, ►



Researchers at New York University's medical school battling to keep samples cold after the flood.

► gave two updates as Richard Cohen, vice-president of facilities management, dealt with a seemingly endless list of logistic issues. The lifts were out of action, so my colleagues and I carried dry ice and liquid nitrogen up five dark flights of stairs to our labs, guided only by the feeble light of a few glow sticks on the steps and landings. (We were lucky — some labs were on the 13th floor.)

When I arrived, the lab was dark and quiet. The silent refrigerators and freezers held thousands of dollars' worth of kits, antibodies, serum and other lab tools that were slowly thawing, now useless.

As quickly as possible, I gathered everyone in my lab and said: "Tell me the most important experiments you need to do, the ones you planned to publish within the next three years. Tell me how you plan to get the animals, breed them and conduct the experiments. Let's shorten the time between today and getting back on our feet."

To continue her experiments, graduate student Sebnem Tuncdemir went to the Salk Institute for Biological Studies in La Jolla, California, to collaborate with neurobiologists Edward Callaway and Martyn Goulding. Others went to labs around New York, at institutions such as Cornell University and the Memorial Sloan-Kettering Cancer Center (MSKCC). It is hard to express my gratitude for the generosity and acts of kindness shown by the scientists who opened up their labs to us. We found ways to make the best use of the downtime — doing data analysis, writing papers, planning a thesis. Fortunately, my 14-person lab was largely made up of senior people who were finishing papers and looking for jobs, and those who weren't yet in the thick of their projects.

In the first few weeks, we lived without the normal channels of communication.

The university servers were down, so we couldn't send e-mails. Mobile phones were mostly down, too. People who could get onto the Internet communicated using personal e-mail accounts. Facebook went from social network to communication tool and came in handy to send messages to each other. Communication was not the only problem — around one-third of the researchers in our lab didn't have power at home, meaning no warm beds or showers for two to three weeks. Yet most of them still came to work. Amazingly, given the damage, within three weeks power was restored in the lab.

RODENT RESCUE

Over the years, I have sent a sample of each of my transgenic mice to my collaborators so that they could pursue similar work, which I felt was my duty as a publicly funded scientist. I never thought it would one day save my lab. Researchers around the world could now send me back my own mice and offered others of their own — even compound strains carrying multiple alleles. I received more than 150 e-mails offering help in the first week. Six months later, I've regained about 35% of what I lost.

Even though I am slowly re-acquiring my strains, researchers have often bred them with others, meaning that we must breed out the traits we don't want. Sometimes we start from scratch, which also takes time — to get four different genetic traits requires four rounds of breeding. Given that each breeding takes two months, and efficiency is about 50%, that translates to 16 months. We will be lucky to have rebuilt our colony within two years.

Furthermore, I have had to submit at least 50 new contracts that enable institutions to exchange patented alleles (material transfer

agreements) because the original contracts expire after three years. This was particularly frustrating. At times, an off-site location would have a mouse ready for me, but we could not get it shipped because of the legal issues.

Then there is the issue of where to keep the new mice. Because of the damage, we have subcontracted space at the MSKCC, and are housing some animals at commercial suppliers, such as the Jackson Laboratory in Bar Harbor, Maine. We will eventually move the colony to the third floor of the NYU science building, but it won't be ready for two years.

TIME TO REFLECT

Throughout this experience, I've had to adjust my expectations for my lab. Not every project can be delayed for six months and survive — in some cases, our competitive advantage has disappeared, so we have had to let those projects go. This might mean that we will get scooped on data that we would have published first had the hurricane not happened. But it has been liberating to stop running the race of competitive science and focus on where we are still ahead of the curve. The US National Institutes of Health is letting me rewrite some of the aims in my ongoing grants, so that I can use the money to pursue new projects.

There are other perverse upsides to this otherwise awful experience. We are much more prepared to handle a similar situation. I'm going to sit down with my lab and develop an emergency-response plan, in case the unexpected occurs again. I will make contingency plans that enable us to access our e-mails, research data and other information even if our server is down for two weeks.

I like to think that the hurricane has also helped my students in some ways, even though it has been frustrating and heartbreaking, and has set back or ended important projects. There were many times when the students needed to act before getting my approval, simply because we couldn't communicate, and they ultimately made the right decisions. The experience taught them how to be free agents; they are more responsible now for their science.

And there was good news even on the darkest of days. A week after the flood, when workers accessed the soaked room of the former mouse colony, they found something unexpected. Before the hurricane, Jennifer Pullium, director of laboratory animal resources, had asked her staff to move some mice to the highest racks, in case the unthinkable happened. When it did, the rising salt water came within inches of their cages. Against all the odds, they had survived. ■

Gordon Fishell is associate director of the Neuroscience Institute at New York University, New York 10016, USA.
e-mail: fisheg01@nyumc.org

SPRING BOOKS



COGNITIVE SCIENCE

Mind as mirror

Philip Ball gets under the skin of a treatise on the brain as an analogy machine.

I finished this review and stored the file in the 'Nature' folder on my desktop, then e-mailed it to the editor. Or did I? A file, after all, was once a sheaf of papers, and a folder a cardboard sleeve for holding them. A desktop was wooden, and mail needed a stamp (no, it needed a little piece of adhesive paper). But all I did was use an interfacing device (named for the most superficial resemblance to a rodent) to rearrange the settings of some microprocessor circuits.

To see that almost everything we say and do refers by analogy to other things we or others have once said or done — which is the main point of *Surfaces and Essences* — there is no better illustration than our computer software, constructed as a conceptual and visual simulacrum of the offices our parents knew.

Why (science-fiction writers take note) would we invent new categories and labels for things when we can aid comprehension

by borrowing old ones, even if the physical resemblance is negligible? What cognitive scientists Douglas Hofstadter and Emmanuel Sander set out to show is that this sort of elision is not merely a convenience: all our thinking depends on it, from the half-truths of everyday speech ("that always happens to me too!") to the most abstruse of mathematical reasoning. I was convinced, and the ramifications are often thought-provoking. But when authors tell you the same thing, over and over again, for 500 pages, perhaps you'll believe it whether it is true or not.

Hofstadter is famous for his Pulitzer-prize-winning treatise on how we think, *Gödel, Escher, Bach* (Basic Books, 1979). Fans of that dazzling performance might find this book surprisingly sober, but it is also lucid

ILLUSTRATION BY ALEX ROBBINS



**Surfaces and
Essences: Analogy
as the Fuel and
Fire of Thinking**

DOUGLAS
HOFSTADTER &
EMMANUEL SANDER
Basic Books: 2013.
592 pp. \$35

The authors focus most on the use of analogy in language. Moving steadily from words to phrases and narratives, they show just how deeply embedded is our tendency to generalize, compare, categorize and forge links. Individual examples seem trivial until you realize their ubiquity: tables have legs, melodies are haunting, time is discussed in spatial terms, and idioms are invariably analogical, if you get my drift. Thus the lexical precision on which dictionaries seem to insist is illusory — words are always standing in for other words, their boundaries malleable. This flexibility extends to our actions: we see that a spoon can serve as a knife when no knife is available. (Indeed, the spoon then becomes a knife — objects may be fixed, but their labels aren't.)

These arguments can be carried too far. Is to extrapolate to make an analogy, expecting the future to be like the past? Is a Freudian slip an analogy, or mere crosstalk of neural circuits? Is convention an analogy (why don't we write $mc^2 = E$)? Can we, in fact, turn any mental process into an analogy, by that very process of analogy? These are not rhetorical questions: one might, in principle, examine whether the same neural circuitry is involved in each case, for example. But a lack of interest in a neuroscientific examination of the authors' idea is one of the book's irksome lacunae.

In fact, this intriguing, frustrating book seems to exist almost in an intellectual vacuum. Unless one combs through the bibliography, one could mistakenly imagine that it is the first attempt to explore the idea of analogy and metaphor in linguistics, overlooking the work of Raymond Gibbs, Andrew Ortony, Esa Itkonen and many others. And one is forced to take an awful lot on trust. When, for example, Hofstadter and Sander describe the evolution of the concept of 'mother' in the mind of a child as he or she learns to generalize from experience, they offer a plausible story, but no empirical evidence for the developmental pathway they describe.

Neither is there any real explanation of why we think this way. Isn't it perhaps, in part, a way of minimizing the mental

resources we need to engage in a situation, to avoid having to start from scratch with every unfamiliar encounter, object or perspective? Is it an adaptive technique for making predictions? Are mirror neurons part of a built-in cognitive apparatus for analogizing ourselves into others' shoes?

The lack of historical perspective is also a problem; it is as if people always thought as they do now. Analogy was arguably all we once had for navigating experience, for example in the Neoplatonic idea of correspondences, "As above, so below." This "just as ... so ..." thinking remains at the root of pseudoscience as well as science: the Moon influences the tides, so why not our body fluids? In which case, how do we distinguish between good and bad analogies?

There are gems of insight in *Surfaces and Essences*, but again these are flawed by the authors' relaxed attitude towards evidence. An analysis of Einstein's thought is splendid, explaining what is missing from conventional accounts of the discoveries of light quanta, relativity and mass-energy equivalence — namely, the qualities that distinguish Einstein from his peers. These qualities are convincingly shown to be analogical: Einstein was able to take leaps of faith and make connections that postpone

rigour and are certainly not self-evidently true.

As Hofstadter and Sander show, these leaps were based on a conviction that different areas of physics were comparable. Einstein's intuition, which his friend and biographer Banesh Hoffmann was content to leave ineffable, is here taken apart so that some of the inner workings may be seen. An ability to draw deep analogies, the authors say, left Einstein like J. S. Bach on

hearing a theme: "very quickly able to imagine all of its consequences". All very fine — but such a detailed account must surely be supported by Einstein's own words. Almost none are offered; we get only fragments of Hoffmann's commentary.

Who is this fecund book for? Academic linguists will be irritated by the absence of references to other work. Physical scientists aren't indulged until page 450. General readers may find it a marathon. The concept of the mind as an analogy generator is persuasive — but would have been equally so explicated at half the length. ■

Philip Ball is a freelance science writer based in London.
e-mail: p.ball@btinternet.com

AN ABILITY TO DRAW DEEP ANALOGIES LEFT EINSTEIN LIKE J. S. BACH ON HEARING A THEME.

and, page for page, a delight to read. Whether there is any conceptual continuity between the earlier work and this new vision of cognition is debatable, except perhaps that the delight in puns in *Gödel, Escher, Bach* here becomes an assertion that pretty much all our mental processing depends on them.

Analogies are the bread and butter (there we go again) of the visual, literary and theatrical arts, although the authors seem curiously unconcerned about any of these except poetry. Yet Hofstadter and Sander

are really inverting that usual picture: art is not a producer of analogies, but a product of our analogical brains.

➔ **NATURE.COM**

For Susan
Blackmore on
Hofstadter, see:
go.nature.com/z1zrbg

GASTRONOMY

The kitchen revolution

Michael Pollan's latest book will be eaten up by the conscious consumers he created, says **Nathan Myhrvold**.

Michael Pollan is one of the most influential food writers of recent times, and has secured a position as the conscience of a new movement dedicated to local, sustainably produced cuisine. Given this position, it is a surprising admission that until recently he had little interest or skill in the craft of cooking. *Cooked* is the entertaining story of his journey to learn from a series of master cooks, artisan bakers, cheesemakers and brewers.

Pollan is a wonderful writer and his account is told with great wit and humour, which makes for a very entertaining read. The masters he chose are great characters — both in life, and under Pollan's pen.

Other writers have also sought to document their culinary apprenticeships. But *Cooked* has much higher ambitions. "My wager in *Cooked*," Pollan says, "is that the best way to recover the reality of food, to return it to its proper place in our lives, is by attempting to master the physical processes by which it has traditionally been made." This isn't just a well-told tale of how he came to master those processes, it is a book with a mission: to inspire readers to get into the trenches of their kitchens, and to stop letting other people prepare, process and package their meals. It succeeds in making its case, despite occasional lapses.

Many advocacy-oriented books use a direct argument. You should eat this because it is delicious, or because it is fun to make, or because it is healthier. Although each of these is mentioned in *Cooked*, they are sidelines compared with the main purpose: to



score intellectual and political points.

Politically, a strong anti-corporate theme runs through the book, blaming food companies for making us their "prey" with "edible foodlike substances". Much as I agree with Pollan on the sorry state of what is on supermarket shelves, surely we, the eaters, bear at least some responsibility for what we consume.

Intellectually, Pollan grapples, with varying degrees of success, with a fundamental contradiction. On the one hand, he wants to bring food "back to earth" rather than allow it be "abstracted" from the traditional methods and values, the "labor of human hands" or the "natural world of plants and animals".

For Pollan, food is meant to be grounded in the context of a traditional kitchen or farmyard; that is how it achieves legitimacy. Yet, on the other hand, he abstracts food by pulling it out of the kitchen and into the salon as a prop in his very philosophical arguments. When he mixes quotes from obscure French philosophers with dialogue from barbecue pitmasters, the result ranges from interesting in some passages to unsuccessful in others. The book's sections mirror the ancient taxonomy of the elements — fire, water, air and earth. But what they are really about is barbecue, bread, beer, pickles and cheese. Put in the patois that his informants might use, if the book is about restoring honesty to food, what's up with the highfalutin words?

In discussing the newfound interest in traditional gastronomy, he asks a rhetorical question: "Can authenticity be aware of itself as such and still be authentic?" It's a very perceptive point in an age in which 'authentic' cuisine — like 'real' southern barbecue or artisanal bread baking — has been seized upon, marketed and branded to a high degree, turning its once humble practitioners into television stars. This is Pollan at his best, honouring tradition while gently calling it into question. In the same

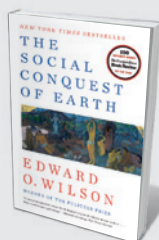


Cooked: A Natural History of Transformation
MICHAEL POLLAN
Penguin: 2013.
480 pp. \$27.95

ILLUSTRATION BY ALEX ROBBINS

**NEW IN
PAPERBACK**

*Highlights of this
season's releases*



The Social Conquest of Earth

Edward O. Wilson (Liveright, 2013; \$17.95)

Distinguished sociobiologist E. O. Wilson asks how social creatures like humans and ants have achieved such evolutionary success. The key, he suggests, is in the way they form communities: with multiple generations, a division of labour and altruistic behaviour. Although Wilson's emphasis on group selection is controversial, this is a masterly amalgam of biology, linguistics, psychology, economics and the arts. (See James H. Fowler's review: *Nature* **484**, 448–449; 2012.)

spirit, I will observe that it is also a question that readers could ask about Pollan's own work, which self-consciously tries to draft on this same authenticity to serve its intellectualism.

Tradition and authenticity are his ideal, but many of his informants aren't as pure as Pollan would like them to be. His barbecue pitmaster uses a proportion of supermarket charcoal, his artisanal baker uses some white flour, his cheesemaking microbiologist nun strikes a nuanced position on raw milk and his pickle guru makes an ersatz kimchi. When this occurs, Pollan wrestles with the issue, sometimes conceding, but often contradicting them or quoting other, more "fundamentalist", sources that call them out for their apostasy.

A scientific perspective on food makes a token appearance, and includes footnotes to papers in scientific journals (including *Nature*). But this is mostly for show; like most books based on traditional cooking, its explanations deviate from scientific accuracy. This book is, at its heart, about what people feel about food, rather than what science has shown to be true.

Pollan's proselytizing that we all ought to cook more can seem a bit strident given that we are living in the golden age of organic, sustainable artisanal local food. Interest in cooking has never been higher (even if many people still don't do it); indeed, that is why Pollan's previous books have been best sellers, as this one is also likely to be. In one passage he marvels that an artisanal baker sells his loaves for only 41 cents more than the giant Hostess Brands sells its Wonder Bread. The unspoken irony is that Hostess itself recently went bankrupt. Times have changed, and many parts of *Cooked* read like a call-to-arms for a revolution that is already well under way, thanks in part to Pollan's previous books. *Cooked* will add to that legacy. ■

Nathan Myhrvold is chief executive and founder of Intellectual Ventures. He is also the creator and co-author of the award-winning books *Modernist Cuisine* and *Modernist Cuisine at Home*.

BIOLOGY

Vive la différence

Suzanne Alonzo relishes a synthesis of the extraordinary variations among males and females of the same species.

Forget men and women being from different planets. In *Odd Couples*, Daphne Fairbairn shows that males and females of many species look almost as if they hail from different galaxies. What is a little friction over whether the toilet seat should be left up or down? You could be a female giant seadevil with a parasitic mate one-fiftieth of your size stuck to you for his entire adult life — or a male garden spider, eaten by your mate after you have broken off your genitals to ensure her fidelity.

Fairbairn, an evolutionary biologist, demonstrates that such differences between the sexes are a fundamental component of biological diversity, affecting everything from an animal's behaviour and appearance to its life expectancy and nervous system. After a general introduction to how this works, Fairbairn spends the bulk of the book on a guided tour of sexual dimorphism in eight carefully selected and researched species, covering two fishes, a bird, a mammal and four diverse invertebrates.

As Fairbairn lucidly explains, the defining distinction between the sexes is that females make eggs and males make sperm. What is harder to understand is how that — along with a species' basic biology and habitat — can drive a cascade of differences in almost every aspect of male and female biology. Whether an organism makes eggs or sperm can affect, for example, the energy it takes to reproduce. This, in turn, affects how much energy each sex has left for growth and survival. Disparities in these, in their turn, alter the body size, habitat use, metabolic rate and reproductive behaviour favoured by Darwinian selection in males versus females. Over time, these effects lead to striking differences in body mass, colour and much more between males and females of the same species. It remains a challenge to understand how these myriad factors interact to shape

the striking differences in what it means, across species, to be male or female.

Fairbairn's tour elucidates these points as it entertains. After first exploring the perhaps more familiar patterns found in mammals and birds (elephant seals and the great bustard, species in which males vastly outweigh, and compete for, females), we encounter much stranger creatures. Take the bone-eating tubeworm: deep below the ocean's surface, harems of dwarf males live within the tube-like home of a single, much larger, female. Even more bizarre are the shell-burrowing barnacles, whose long-lived females weigh 500 times as much as the short-lived males. The males never eat, developing into little more than sperm production and delivery machines on finding a female.

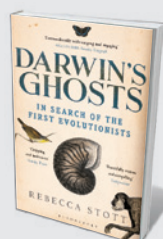
A key message here is that the large, flashy males who fight one another for access to numerous small, coy females — as seen in birds and mammals — are not representative of the predominant pattern. Females are larger in 86% of animal classes with sexual size dimorphism, Fairbairn tells us, and in many species the main challenge

males face is finding a female. Moreover, Fairbairn emphasizes that selection on males and females differs in a multitude of ways, rather than being primarily due to sexual selection on males (namely, competition among males for access to mates or to fertilize eggs). For example, male shell-carrying cichlid fish are much larger than females of the same species not only ▶



Odd Couples: Extraordinary Differences Between the Sexes in the Animal Kingdom

DAPHNE J. FAIRBAIRN
Princeton University Press: 2013. 312 pp.
\$27.95, £19.95



Darwin's Ghosts: In Search of the First Evolutionists

Rebecca Stott (Bloomsbury, 2013; £8.99)
Science historian Rebecca Stott probes the intellectual origins of the theory of natural selection, showing that Charles Darwin stood on the shoulders of giants, from Aristotle to Jean-Baptiste Lamarck. (See Andrew Berry's review: *Nature* **485**, 171–172; 2012.)



The Spark of Life: Electricity in the Human Body

Frances Ashcroft (Penguin, 2013; £9.99)
As you read this, ion channels regulate the electrical activity in your neurons and muscle cells. Physiologist Frances Ashcroft offers a brilliant treatment of the 'body electric', mixing research, science history and personal stories.

► because reproductive competition among males for territories favours size — but also because selection favours females small enough to fit inside a shell to care for their young.

Finally, although the possible biological origins of human sex differences continue to fascinate, human sexual dimorphism is really not that striking. Men and women are boringly similar in size compared with other primates, and obviously outclassed in the oddity stakes by the other species highlighted here.

Fairbairn has simplified some material and left certain complexities out. For instance, there is nothing on the recent research documenting striking differences between the sexes in gene expression, affecting everything from early development to social behaviour, and little on the fact that we have only just begun to understand how a single genome can produce such diverse forms. But *Odd Couples* is a pleasure to read. There is humour (including an eye-rolling joke or two), but no reliance on the anthropomorphic cuteness so common in popular books on animal behaviour — especially sexual behaviour. There are certainly moments where the author ‘geeks out’ on the details, and this is part of the appeal. You walk away from this book with a deeper understanding of both these creatures and a biologist’s mind.

I am inevitably biased in favour of Fairbairn’s theme, having spent my working life trying to understand the amazing diversity of reproductive behaviours. Even so, I found reading the book like taking a holiday in a foreign land with an enthusiastic and expert guide. You will come back with good stories, and a new appreciation of the amazing diversity of life on Earth and the forces shaping it. You may even find your perspective on bigger questions shifting.

As Fairbairn concludes: “The enduring message from all of this is that there is clearly no one way of being a male or a female.” When it comes to sex roles, all bets are off in the animal kingdom. ■

Suzanne Alonzo is an evolutionary biologist in the Department of Ecology and Evolutionary Biology at Yale University in New Haven, Connecticut.
e-mail: suzanne.alonzo@yale.edu



GEOLOGY

Written in stone

Ted Nield relishes a deft tracing of the relationship between the rise of geology and the novel in the turbulent nineteenth century.

When we imaginatively recreate the past, we enter a dangerous landscape: we may find ourselves needing a philosophical map. Things become even more treacherous when trying to recreate the ways our ancestors looked back at history. This entails deciphering a palimpsest. Its cartographic vagaries may further distort our hindsight. Adelene Buckland attempts just such a recreation in her book *Novel Science*.

Buckland tries to get inside the heads of the Britons who were writing into existence a scientific geology while developing

a great literary form: the nineteenth-century novel. She succeeds triumphantly.

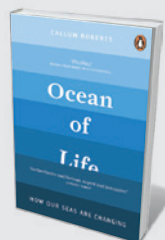
Like their descendants today, the groups driving these two grand projects were not much separated from each other in the late eighteenth and early nineteenth centuries. Victorian geologists, and Charles Lyell in particular, were deeply concerned with

evolving appropriate literary and visual forms that would convey their geological discoveries. The creative

NATURE.COM

For more on Charles Dickens and science, see go.nature.com/79ckns

THOMAS JACKSON/STONE/GETTY



Ocean of Life: How Our Seas Are Changing

Callum Roberts (Penguin, 2013; £10.99)
Overfishing, acidification, plastic pollution, biogeographical shifts: marine conservation biologist Callum Roberts lucidly lays out the range of issues affecting the world’s oceans. A sobering look at Earth’s biggest biosphere. (See Stephen R. Palumbi’s review: *Nature* **484**, 445–446; 2012.)



Antarctica: An Intimate Portrait of the World’s Most Mysterious Continent

Gabrielle Walker (Bloomsbury, 2013; £8.99)
Science writer Gabrielle Walker unveils Earth’s southernmost ‘wild lab’ in this vivid and accessible mix of researchers’ stories and environmental writing. (See Francis Halzen’s review: *Nature* **483**, 272–273; 2012.)

act of writing was, for them, as essential a part of scientific practice as any other, and they looked to contemporary writers of fiction for models. Meanwhile, those novelists — beginning with Walter Scott, and later including the likes of George Eliot, Charles Kingsley and even Charles Dickens — drew from the new science of geology and the awareness of deep time that it brought into popular consciousness. They found a new profundity with which to disturb and enrich their narratives.

The evolution of these two fields, geology and literature, mirrored and drove each other. The scientists sought to develop rigour, the novelists to achieve seriousness. 'Romance', in both cases and senses, was the enemy.

Buckland begins by taking us through the emergence of geology from its highly speculative, theoretical roots. In the early to mid-eighteenth century, speculation about Earth's structure and history was the preserve of *Weltall* theorists — system-builders who focused on how the cosmos began. They devised all-encompassing cosmogonies, then cherry-picked their evidence to suit. Even the Scottish geologist James Hutton, whose *Theory of the Earth* (first made public in 1785) ushered in a properly constrained, scientific approach to the rock record, sat within this tradition. But Hutton introduced — and Lyell firmly established — a key principle that University of Cambridge don William Whewell termed 'uniformitarianism' in the 1830s. This doctrine, which holds that all interpretation of the past must refer to processes that can be seen operating on Earth today, remains the central concept that makes geology 'scientific'.

Within uniformity, however, questions remained — even into our own times. Did today's processes always operate at today's rates? Is the tiny snapshot of human experience an adequate sample of Earth history? And

does the occasional rare event leave more of a trace in the record than the long ages that pass in between?

Lyell adhered to an overly strict constancy of rate for Earth processes — perhaps because, as Buckland reminds us, he

THE DOCTRINE THAT ALL INTERPRETATION OF THE PAST MUST REFER TO PROCESSES THAT CAN BE SEEN OPERATING ON EARTH TODAY MAKES GEOLOGY 'SCIENTIFIC'.

trained as a lawyer. Using his chief skill of rhetoric, he sought to establish that, on an Earth of extreme age, everyday processes would efface any occasional catastrophe. For Lyell, gradualism was all. Another crucial turning point on the road to rigour and respectability was the foundation of The Geological Society of London in 1807, in whose hallowed halls I work. The society set itself against all theorizing in favour of information-gathering.

But the society's literate builders of geology, such as Lyell, William Buckland and William Conybeare, fretted that their science might be embodied in or even traduced by literary forms that militated against the quest for academic dignity. Their loathing of 'theory' led them to suspect any reliance on its narrative analogue, 'plot' — with its emphasis on causality and motive. They reviled popularizers such as Robert Chambers — revealed as the author of the scandalous 1844 book *Vestiges of the Natural History of Creation* only after his death — who succumbed to such literary devices. (Some things don't change much.)

Wishing to purge their science of romance, they sought a drier narrative

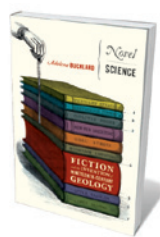
approach. This could have endangered their mass appeal. Happily, it didn't. Lyell and his peers each assumed the role of the wandering romantic, allowing a public fascinated by their discoveries to picture the heroic geologist — such as the weatherbeaten Adam Sedgwick pausing atop Glyder Fawr, one of Wales's highest mountains, like some human embodiment of painter Edwin Landseer's *The Stag at Bay*.

Meanwhile, contemporary novelists were inserting discursive philosophical elements into their writing. As Buckland argues, Scott did the most to reinvent the novel for his contemporaries as a credible literary form fit for gentlemen to read, as well as ladies. Scott, followed by Elizabeth Gaskell, Eliot, Kingsley and others, distanced their art from the yarn-spinning romancers of yore, such as Laurence Sterne, who cleaved more to the ancient traditions of Miguel de Cervantes and François Rabelais.

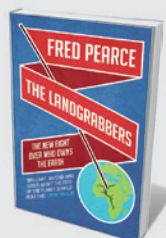
As both groups strove for realism, geologists discovered Scott, and he them. Buckland's book is the story of how they, and successive generations of geologists and novelists, helped one another to write the past into existence. It culminates, for me, in the work of geologist-novelist Kingsley, who even seems to have striven for the fusion of story-line and stratigraphy. Buckland will send you scouring the second-hand bookshops for long-forgotten works.

The relationship between science and literature has proved to be a rich seam of inquiry since 1983, when Gillian Beer produced her seminal book *Darwin's Plots* (Cambridge University Press). In the intervening decades, Earth scientists, with their strong historical bent, have worked with science historians and literary critics to create today's vibrant, culturally integrated field. A few inconsequential slips apart (neither William Buckland nor Conybeare were among the 13 founders of the Geological Society of London), Buckland meets this multidisciplinary challenge well in *Novel Science*. ■

Ted Nield is editor of *Geoscientist*, the Geological Society of London's monthly fellowship magazine. His next book, *The Forgotten Land*, is expected early next year. e-mail: ted.nield@geolsoc.org.uk

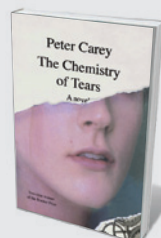


Novel Science: Fiction and the Invention of Nineteenth-Century Geology
ADELE BUCKLAND
University of Chicago Press: 2013. 384 pp. £29, \$45



The Landgrabbers: The New Fight Over Who Owns the Earth

Fred Pearce (Eden Project Books, 2013; £9.99) Delving into the recent 'land grabs' in developing countries, science journalist Fred Pearce mulls over solutions, such as including African smallholders in the global agricultural economy. (See Wendy Wolford's review: *Nature* **485**, 442–443; 2012.)



The Chemistry of Tears: A Novel
Peter Carey (Vintage, 2013; \$15)

The history of science and engineering flavours this moving novel centring on a nineteenth-century automaton. Peter Carey's meditation on time and early 'artificial life' raises questions about what it means to be human. (See Minsoo Kang's review: *Nature* **484**, 451–452; 2012.)

PHYSICS

Clockwork cosmos

Pedro Ferreira ponders a vision of the Universe in which time is paramount.

Theoretical physicist Lee Smolin's recent books have been about crises in physics so catastrophic that physicists need to completely rethink their methods. In his 2006 book, *The Trouble with Physics* (Houghton Mifflin), he stated controversially that a cabal of researchers working on what he thought was a moribund theory of fundamental physics — string theory — was preventing a new generation of clever young thinkers from working on other, rival theories. Through his brilliant writing and articulate arguments, readers took him seriously. One string theorist told me that he struggled to convince non-physicists that he wasn't a charlatan after the publication of Smolin's book.

Now, in *Time Reborn*, Smolin attempts to chip away at basic theories of modern physics. He makes the case that by doing away with time, existing theories are missing a trick. He uses the orbits of planets in the Solar System as an example: each orbit is an ellipse existing in three dimensions. A planet will lie, at some moment, on a point along that track. But its motion can be described without knowing what happens at that particular moment, or at any other. Newtonian physics is essentially timeless.

According to Smolin, our picture of a timeless Universe stems from the assumption that all modern physics — quantum as well as classical — is predictive. How a system evolves is entirely encoded in the starting set of 'initial conditions' and their transformation according to the laws of physics. Evolution in time is



Time Reborn: From the Crisis in Physics to the Future of the Universe
LEE SMOLIN
Houghton Mifflin Harcourt:
2013. 352 pp. \$28, £20

secondary, a by-product of the theory. This bothers Smolin. A timeless view of reality is, he says repeatedly, incomplete (where do the initial conditions or laws come from?) and, simply, "wrong". He believes that a

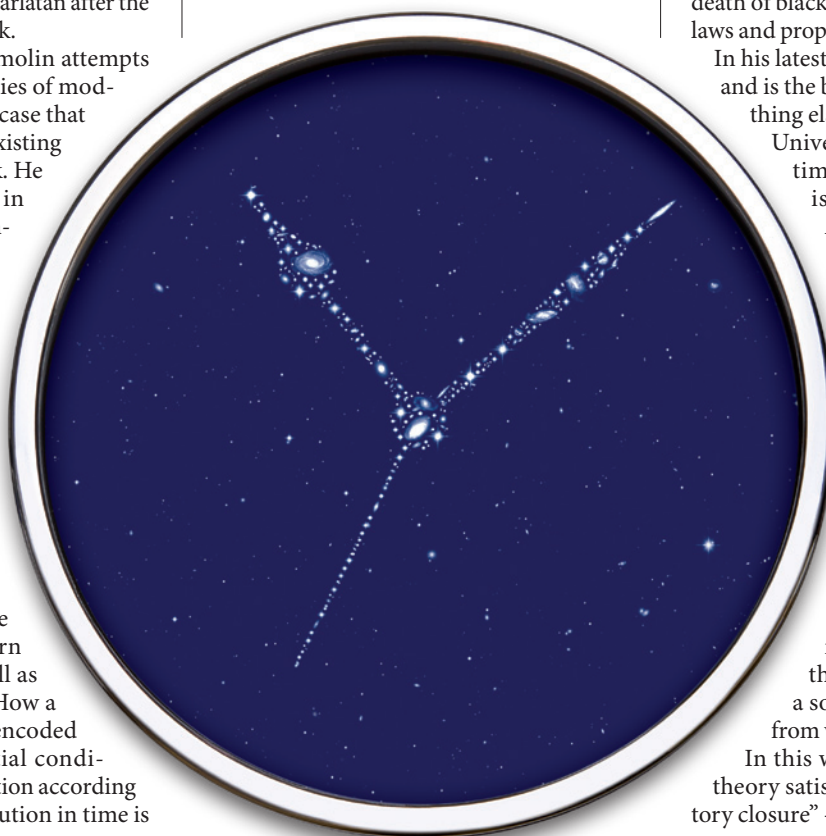
better description of time lies at the heart of some of the big questions, such as the marriage of quantum physics and general relativity.

Smolin sketches an alternative path for modern physics. Inspired by the ideas of Brazilian philosopher and political theorist, Roberto Mangabeira Unger, who argues that social structures emerge without an underlying natural order or guiding principle, Smolin develops some of the ideas behind his first book, *The Life of the Cosmos* (Oxford University Press, 1997). In it, he argued that the Universe evolved through natural selection, mediated by the birth and death of black holes, to give us the physical laws and properties we measure today.

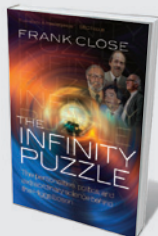
In his latest vision, time reigns supreme and is the backbone from which everything else emerges. Each state of the Universe pops up somewhere in time, from what the Universe is made of to what it does.

A prime example is space, which — echoing some of the ideas put forward by different schools of quantum gravity — emerges not as a fundamental entity, but as a tapestry of connections between events happening over time. More importantly for Smolin, none of the laws or principles that we have discovered over the centuries constitute the bedrock of physics, nor are any perennial. On the contrary, they emerge in a somewhat unpredictable way from what is going on at each time.

In this way, he says, his embryonic theory satisfies a "principle of explanatory closure" — there is no need to invoke

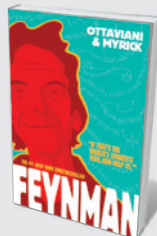


ALEX ROBBINS/SHUTTERSTOCK



The Infinity Puzzle: The Personalities, Politics, and Extraordinary Science Behind the Higgs Boson

Frank Close (Oxford Univ. Press, 2013; £10.99)
Particle physicist Frank Close pins down the elusive Higgs boson in this account of the search that led up to its 2012 discovery. With a Nobel prize in the offing, the vexed question of credit adds edge. (See Edwin Cartledge's review: *Nature* **478**, 315–316; 2011.)



Feynman

Jim Ottaviani and Leland Myrick (*First Second*, 2013; \$19.99)

The playful creativity and genius of theoretical physicist Richard Feynman are brilliantly brought to life in Jim Ottaviani's graphic biography, illustrated by Leland Myrick. (See Marc Weidenbaum's Q&A: *Nature* **477**, 32; 2011.)

any external laws or initial conditions.

It is a tall order, and if Smolin's theory is to work, then all the great experimental discoveries in physics — from elliptical planetary orbits to the Higgs boson — need to be incorporated. Hallowed theories such as quantum physics and relativity must be dismantled and some radically new way of explaining how the Universe evolves must come into play. Smolin shies away from actually telling us what that new way is, because he doesn't seem to know himself. All he can do is to explain how different his theory must be from everything we have done before.

To explain why anything can be predicted at all in such a lawless Universe, Smolin invokes reproducibility: if a physical process has happened in a certain way before, it will happen in the same way again. We can predict what will happen if we have some familiarity. But, Smolin notes, there will be situations that we have never seen before, in which it will be impossible to predict the outcome.

Writing a book is a well-worn way of presenting a provocative theory that is still in its infancy. Smolin, a respected physicist with a track record of best-sellers, has a privileged platform for promoting his ideas, similar to Arthur Eddington, Erwin Schrödinger or Fred Hoyle before him. Books can, however, feel reckless without the filter of the (albeit flawed) peer-review process.

Yet I enjoyed *Time Reborn*. Smolin is an excellent writer, a creative thinker and is ecumenical in the way he covers so many different branches of thought. Even as I mentally argued with this book, I kept on ploughing through to see how Smolin dealt with the objections. I would love to sit down with him over a drink and debate the ins and outs of his theory. And that is how this book should be read: as an account that makes you ask questions. ■

Pedro Ferreira is professor of astrophysics at the University of Oxford, UK.
e-mail: p.ferreira1@physics.ox.ac.uk



NEUROSCIENCE

Drugs to build a better brain

Anjan Chatterjee probes a cognitive-enhancement primer.

Decisions can be as trivial as which coffee to order or which wine to buy, or as consequential as who to marry or which job to accept. Yet even the most profound choices are rarely made on strictly logical grounds. We don't weigh up pros and cons and dispassionately pick the best course of action. Our emotions and attitude to risk, how a situation is framed and the time available all influence our final choices.

In *Bad Moves*, Barbara J. Sahakian and Jamie Nicole Labuzetta lay out the neuroscience of how people make decisions and the ethical quandaries that accompany the use of drugs to enhance cognition. Their slim book is admirable in reviewing these important topics, but it does little to explore the wider view of how emotions can be regulated by drugs.

Sahakian, well known for her research on the neuropsychology of affective and cognitive systems, and neurologist Labuzetta use people with dementia, depression, mania and phobias, who tend to make poor

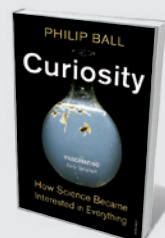


Bad Moves: How Decision Making goes Wrong, and the Ethics of Smart Drugs
BARBARA J. SAHAKIAN
AND JAMIE NICOLE LABUZETTA
Oxford University Press: 2013. 192 pp.
£14.99

decisions, as exaggerated examples of how we can all err. Abnormal functioning of the frontal lobes and deep limbic structures in the brains of people with these disorders disrupts their emotional control and thus decision-making ability.

After discussing decision-making processes in the brain, Sahakian and Labuzetta explore cognitive enhancers. They focus on cholinesterase inhibitors and stimulant medications that can improve memory, sharpen attention and boost concentration. Such 'smart drugs' raise an ethical question: if drugs developed to treat people with

ILLUSTRATION BY ALEX ROBBINS



Curiosity: How Science Became Interested in Everything

Philip Ball (Vintage, 2013; £9.99)
Humanity's burning urge for knowledge drives science. Philip Ball's scintillating history of curiosity brims with treats — such as seventeenth-century philosopher Francis Bacon's use of a Pan myth as an allegory for the quest to learn from nature.



Genentech: The Beginnings of Biotech
Sally Smith Hughes (Univ. Chicago Press, 2013; \$16)

The history of Genentech, the company that kick-started the biotech industry, is compellingly told by Sally Smith Hughes. Studded with in-depth portraits of its pioneers. (See Linnaea Ostroff's review: *Nature* **478**, 456; 2011.)

► cognitive disorders can also make people with healthy brains smarter, should we use them?

There is no simple answer. Smart drugs can make us more efficient and productive, which may be a good thing for society. But there are many reasons to be cautious. The long-term safety of ingesting these drugs is not fully known, although stimulants can be addictive. Easy rewards from these medications undermine the value of hard work and threaten our ideas of authenticity. And the availability of such drugs could compromise our liberties.

We could feel compelled to use drugs of this kind if all those around us are taking them and appear more productive. We might even insist that some people, such as commercial pilots and medical residents, take cognitive enhancers. And variations in access to smart drugs could raise concerns of fairness and justice, particularly if the advantages they confer are available disproportionately to the rich.

Although the book's themes are timely, the link between them is not transparent. After the authors make the convincing case that emotional dysregulation can cause us to choose badly, I expected a discussion about our ability to regulate emotions chemically. Surprisingly, the authors make no mention of antidepressants, anxiolytics and mood stabilizers, and the ethics of their use in healthy people. As a result, Sahakian and Labuzetta's diagnosis of the emotional source of bad decisions is disconnected from potential interventions.

Nonetheless, *Bad Moves* offers a good introduction to issues that affect us all. As the authors astutely point out, academics are not the final arbiters of the ethics of cognitive enhancement — these are societal concerns. With this accessible primer, full of medical anecdotes and clear explanations, Sahakian and Labuzetta prepare the public for an informed discussion about the role of drugs in our society. ■

Anjan Chatterjee is professor of neurology at the University of Pennsylvania in Philadelphia.
e-mail: anjan@mail.med.upenn.edu



CULTURE

Of Genesis and genetics

Tim Radford revels in a masterly take on science invoked by the Bible.

The *Serpent's Promise* is a believer's book. It expresses belief in the power of language, imagination, scholarship, high art, enduring myth, tribal tradition, unforgettable poetry, irrational vision and inspired insight. If you wanted to find all of these things between just one set of covers, you might pick up the Authorized Version of the Bible; but this is not a book by somebody who believes in God. It is a book by the distinguished geneticist, broadcaster, lecturer, writer and Welshman Steve Jones, who has a sharp awareness of moral imperative

and a warm feeling for those Joneses before him who invoked the bread of heaven and yearned to be safe on Canaan's side. It is the ambivalence at the heart of this book which makes it so hugely enjoyable and, perhaps, so important.

Jones' story is not of the science of the Bible, but of the science invoked by the Bible. The Good Book (his words, his capitals), he says, was always more of a guide

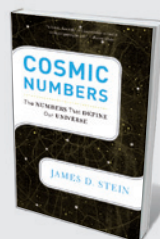
ILLUSTRATION BY ALEX ROBBINS

NATURE.COM
For Mark Pagel on
Steve Jones's *Almost
Like a Whale* see:
go.nature.com/ca6kzj



The Signal and the Noise: The Art and Science of Prediction

Nate Silver (Penguin, 2013; £8.99)
Statistician Nate Silver reveals how 'noise', a random component of data, often clogs up the complex process of forecasting. Silver makes a convincing case for a Bayesian approach (See Paul Ormerod's review: *Nature* **489**, 501; 2012.)



Cosmic Numbers: The Numbers that Define Our Universe

James D. Stein (Basic Books, 2013, \$15.99)
Key numbers in physics, chemistry and astronomy star in this mathematical history. James D. Stein captures ideas from luminaries such as Isaac Newton and Johannes Kepler to characterize these 'universal' measurements.

book, “a handbook to comprehend the world ... it sits firmly in the genealogy of ideas. Science is its direct descendant.” In each chapter he takes a text — from Genesis or the Gospel of John, from Ecclesiastes or Matthew, from Exodus, Leviticus, Job and so on — as the starting point for a rationalist sermon on a biblical theme. So Jones uses Genesis 6:4 (“There were giants in the earth in those days”) as a springboard less for talking about Goliath than for using “the power of science to illuminate myth” and for discussing the growth-hormone disorder acromegaly, linked to tumours of the pituitary gland. The long life described in Ecclesiastes 11:8 prompts reflections on insulin, the French paradox (high consumption of saturated fats coupled with low rates of coronary heart disease), the joys of red wine, the connections between sex and death and the enhanced lifespans of castrati.

His choice of stories from the Bible (Noah’s Ark and the flood, Joseph in Egypt and the years of plenty and famine, among others) are no surprise. The delight is in the delivery — often witty and laconic, always generous. He does not waste much energy on the three great mysteries resolved with such confidence in Genesis (“the world’s first biology textbook”): science may never be able to explain why the Universe happened at all, precisely how life began or what exactly turned an omnivorous foraging African bipedal primate

into a creature with a taste for abstract speculation. The reward arrives with all those other Biblical preoccupations — Eden, a homeland, long-lived Methuselah, dietary rules that distinguish one group from another, the treatment of leprosy, the emeralds or swellings with which

God smote the Philistines, and ancient and modern insurance policies. (“Noah, unlike his feckless fellows,” writes Jones with a characteristic flourish “was seen as a good bet in the eyes of the Lord and quite soon, his policy paid off.”)

He is, of course, terrific on genetics.

THAT IS THE PROBLEM WITH HUMANS. THEY CAN INTELLECTUALLY ENDORSE ONE THING AND STUBBORNLY LOVE ANOTHER.

Jewishness is historically defined by descent, and the Bible is big on begetting. The stories told in human DNA sometimes square with tradition, and sometimes do not. Yes, the human race was all but extinguished — but perhaps more than once. Yes, the mutations in the male Y chromosome point back to a single progenitor in Africa 100,000 years ago. But the mother of all humans — the only one whose daughters all had daughters — lived in Africa 200,000 years ago. Adam and Eve can never have met, “let alone have committed the first and perhaps least original of all sins”.

About half of all the Ashkenazim, the biggest group of Jews, share descent from just four women (the number of women who survived on the Ark, Jones teasingly reminds us). Half of all Russian males have a Y chromosome linked to the historical Arya people of Iran. But this is not the case in Germany — Teutonic purists of the early twentieth century who claimed Aryan supremacy in fact shared their chromosomes with people in the Middle East. They had on average a closer tie with the Jewish men they despised than with the Arya. Almost all native Britons can trace descent from a single anonymous individual who lived around the thirteenth century. The most recent universal common ancestor for the entire planet dwelt about 100 generations

ago in the Bronze Age, perhaps around the time of the destruction of Solomon’s Temple in Jerusalem in 600 BC. As we count back through the generations, our ancestors multiply. But populations were smaller, so we begin to share forebears. We have roots in common, says Jones: “Ancestry is a forest not of pines but of mangroves.”

In 1999, in *Almost Like a Whale* (Doubleday), Jones updated Darwin, starting each chapter with Darwin’s own words: hardly an impertinence, given that every evolutionary biologist updates Darwin. *The Serpent’s Promise* cannot advance divine revelation, but it offers a new context for old myths. It is of course

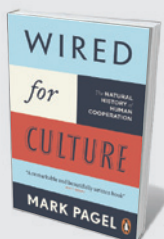
superbly written by someone who quotes historian Edward Gibbon, Marxes Karl and Groucho, Mark Twain, James Boswell and Giovanni Boccaccio, and gourmet Jean Anthelme Brillat-Savarin with the casual ease of an omnivorous reader. This book is not an overt condemnation of religious belief: skilfully, it selects stories that have informed Western culture for 2,000 years to illuminate modern research, and Jones ends with an envoi on behalf of a future enriched by “an objective and unambiguous culture whose logic, language and practices are permanent and universal. It is called science.”

I don’t think even Jones believes that things are going to work out that way, if only because he also begins each chapter, and the book, with illustrations by William Blake, “who demonstrates, better than almost anyone else, the power of sacred imagery to move even those who do not share his convictions”. That is the problem with humans. They can intellectually endorse one thing and stubbornly love another, which is why *The Serpent’s Promise* is more than just another science book, and all the more humane for its wider dimension. ■

Tim Radford is a former science editor of *The Guardian*, and author of *The Address Book: Our Place in the Scheme of Things*. e-mail: radford.tim@gmail.com

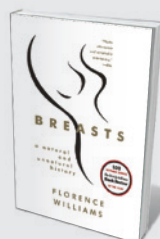


The Serpent’s Promise: The Bible Retold As Science
STEVE JONES
Little, Brown: 2013.
448 pp. £25



Wired for Culture: The Natural History of Human Cooperation

Mark Pagel (Penguin, 2013; £9.99)
Culture has shaped us, besting even genes, says evolutionary biologist Mark Pagel. Full of gems, such as the similarities between ‘tree’ diagrams for languages and for related species. (See Peter Richerson’s review: *Nature* **482**, 304–305; 2012.)



Breasts: A Natural and Unnatural History

Florence Williams (W. W. Norton, 2013; \$15.95)
In this meticulously researched environmental history, Florence Williams covers the human breast from puberty to menopause and beyond. Fascinating, from its unique development to the toxins lurking in breast milk. (See Josie Glausiusz’s review: *Nature* **485**, 306–307; 2012.)

Correspondence

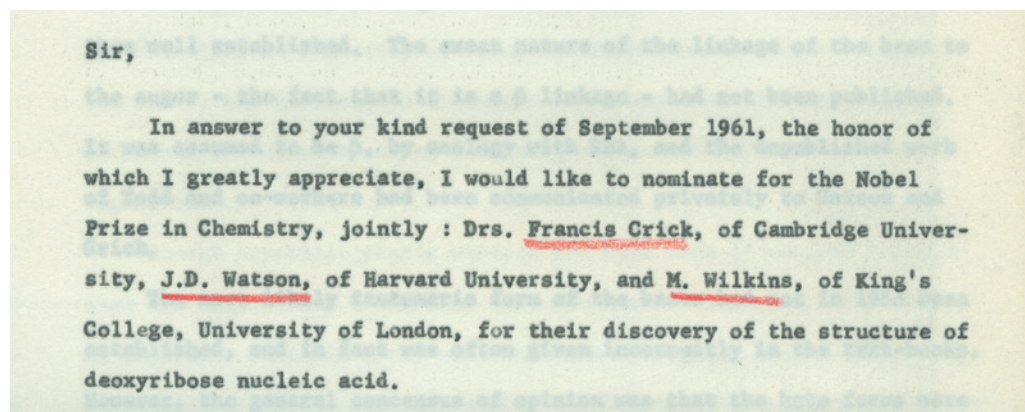
DNA: archives reveal Nobel nominations

Recently released letters shed light on the Nobel prize nominations for the discovery of the DNA double helix 60 years ago.

On 31 December 1961, Francis Crick sent Jacques Monod, at Monod's request, a nine-page account of the discovery of the structure of DNA (see D. T. Zallen *Nature* **425**, 15; 2003). Crick laid out what was known before work on the structure began in 1950, detailed his and James Watson's contributions and summarized work confirming that their model was correct. Crick wrote, "I hope it [the account] is not far from the sort of thing you wanted. It really is most kind of you to take all this trouble on our behalf" (source: Wellcome Library, London).

This has been taken to mean that Monod was preparing to nominate Watson and Crick for the Nobel Prize in Physiology or Medicine, which they won in 1962 with Maurice Wilkins. Watson, in his 2007 book *Avoid Boring People* (Knopf) wrote: "Jacques Monod [...] could not keep secret from Francis Crick that a member of the Karolinska Institutet in Stockholm had asked him to nominate us in January for the 1962 Nobel Prize in Physiology or Medicine."

We were therefore surprised not to find Monod's nomination letter among those released by the Nobel Committee for Physiology or Medicine. We found it instead in the archives of the Pasteur Institute in Paris, and, contrary to received wisdom, the nomination was for the prize in chemistry



(see letter, pictured). In the event, the 1962 chemistry prize went to Max Perutz and John Kendrew for their determination of the structures of haemoglobin and myoglobin.

The fact that the double helix was the subject of nominations for both prizes must have presented a dilemma for the two committees. This was highlighted by a letter from Nobel laureate George Beadle (who had won the medicine prize himself in 1958) nominating Crick, Watson and Wilkins for the 1961 prize. After agreeing that the structure deserved recognition through the chemistry prize, he went on: "But I also feel — and most strongly — that it is so important for biology that it should be recognized by the Prize in Physiol. & Med. if the chemists do not do so." Perhaps, as science historian Horace Judson put it, "The Nobel committees, with a lightness of touch they had not been known to possess, had gotten together to give prizes for the two discoveries [...] made in the Cavendish Laboratory in 1953." (H. F. Judson *The Eighth Day of*

Creation CSHL Press, 1996).

The earliest nomination mentioning the DNA structure was from British virologist Michael Stoker, who recommended Crick and Watson for the 1960 physiology or medicine prize. This was followed by three nominations for the 1961 prize and two for the 1962 prize (see table). The first chemistry nominations (from Jacques Monod, Peter Campbell, William Stein, Harold Urey, John Cockroft and Stanford Moore) were for the 1962 prize. (Information from the Nobel Archives, The Royal Swedish Academy of Sciences.)

Crick's letter to Monod acknowledges the importance of Rosalind Franklin's X-ray data for certain features of the structure. Franklin died in 1958 and, because the Nobel prize is not awarded posthumously, she could not have been considered in 1962, nor indeed at the time of any of the earlier nominations. **Alexander Gann, Jan A. Witkowski** *Cold Spring Harbor Laboratory, New York, USA.* witkowsk@cshl.edu

DNA: twin strands solved the structure

Today is the 60th anniversary of the publication in *Nature* of three papers on the structure of DNA, by James Watson and Francis Crick, and by teams led by my late father, Maurice Wilkins, and Rosalind Franklin (*Nature* **171**, 737–738; 738–740 and 740–741; 1953). It is easy to forget that, in April 1953, the few scientists who had even heard of DNA mostly dismissed it as unimportant.

My father wrote to Watson and Crick at the time: "There is no good grouching — I think it's a very exciting notion and who the hell got it isn't what matters." I doubt that anyone connected with that letter would have believed how much "grouching" about 'winners' and 'losers' the next 60 years would bring.

The structure of the DNA double helix emerged from the twin strands of the University of Cambridge's conceptual model and King's College London's experimental rigour. Both contributions were vital to its precision and validation.

The four different figures in the 'race for DNA' shared a common concern about the effect of science, including their own, on humankind. None could have expected that their work would have such an impact. Let's hope the end result of this "very exciting notion", 60 years young, is that we'll all be the winners.

George Wilkins *London, UK.* georgewilkins1@hotmail.co.uk

NOMINATIONS FOR THE NOBEL PRIZE IN PHYSIOLOGY OR MEDICINE

Nominator	Nomination submitted	Prize year	Nominees
Michael Stoker	22 January 1960	1960	Francis Crick and James Watson
George Beadle	19 November 1960	1961	Crick and Watson; also suggested Maurice Wilkins
Albert Szent-Györgyi	6 December 1960	1961	Crick and Watson
Gilbert Mudge	23 February 1961	1961	Crick and Watson
Charles Stuart-Harris	6 November 1961	1962	Crick, Watson and Wilkins
George Beadle	7 November 1961	1962	Crick and Watson; also suggested Wilkins

FORUM Quantum physics

A grip on misbehaviour

Physicists have come up with a way to characterize and command untrusted quantum systems. Two experts discuss the significance of these findings for fundamental science and for practical quantum computation and cryptography. [SEE ARTICLE P.456](#)

THE PAPER IN BRIEF

- To reliably process information using quantum systems, it is pivotal to check whether the systems are truly quantum and behave as instructed.
- In 1969, Clauser, Horne, Shimony and Holt proposed a test, known as the CHSH test, to detect a feature of quantum mechanics

called quantum non-locality.

- Building on this proposal, Reichardt *et al.*¹ (page 456) extend this test to characterize and control the dynamics of a quantum system.
- The result brings physicists closer to the dream of secure quantum cryptography even when using untrusted data-encryption and data-decryption devices.

Quantum black boxes

STEFANO PIRONIO

Characterizing the state and dynamics of an unknown system is a central problem in most scientific activities. It is a complex process that involves acquiring and interpreting data from various instruments, and often relies on a priori models and approximations that might need later validation. What can we say about a system's behaviour if we have only minimal information about it? Reichardt *et al.* consider the extreme case in which a quantum system, when viewed as a black box from the perspective of an external observer, can be probed only through a simple, digital, classical interface: the observer can ask only two questions, for example by pushing button 0 or button 1, and the system can deliver only two answers, 0 or 1, corresponding, for example, to one of two lights flashing or not (Fig. 1).

The observer does not know what the questions mean, that is, which properties are probed, and is ignorant of the process that produces the answers. The system can be queried as many times as desired, but there is no guarantee that it will behave the same way every time. The information that can be obtained is limited, whereas the system and its quantum dynamics could be arbitrarily complicated.

In this simple scenario, obtaining any useful information about the internal workings of the unknown system seems hopeless. Indeed, it is: many different processes can produce the same sequence of answers, and any such

sequence could simply have been produced by a classical computer.

The situation becomes interesting when, as Reichardt *et al.* consider, instead of one there are two such systems, A and B. Then something non-trivial can be said about their joint behaviour by observing possible correlations between the two systems. Suppose, for instance, that when both systems are probed, system B always produces an answer that is correlated to the question that system A is asked: if A's question is 0, then B outputs 0, and if A's question is 1, then B outputs 1. Some kind of interaction between the two boxes is required to produce this pattern of answers. If no interaction was initially apparent, then we have learned something about the joint dynamics of the systems, although we may still be ignorant of the internal workings of each individual system.

In 1964, John Bell discovered² a feature of quantum theory, known as quantum non-locality, according to which certain pairs of quantum systems, although apparently separated and non-interacting, display strong correlations, almost as if they were a single entity. To demonstrate the phenomenon of quantum non-locality experimentally, Clauser, Horne, Shimony and Holt devised a statistical test, the CHSH test, that can detect non-local correlations between two systems without any assumption about their internal working³, as in the simple example of matching 0s and 1s discussed above.

Researchers have since shown that the CHSH test can detect not only non-local correlations between two quantum black boxes but also other physical properties, such as the amount of quantum randomness produced by the boxes⁴ or, in some circumstances,

their joint quantum state⁵. This is possible because quantum theory imposes relationships between non-locality and those other physical features. In their study, Reichardt *et al.* push this line of reasoning further and achieve a technical breakthrough: they show that the presence of a sufficiently high amount of non-locality, as measured by the CHSH test, characterizes (almost) completely the joint state and individual dynamics of the two quantum black boxes.

Furthermore, they demonstrate that the CHSH test can be used as a tool to realize and control arbitrary quantum dynamics with two non-interacting quantum systems, without making any assumptions about their internal structure. These results are not only conceptually fascinating, but, as discussed below, they also have profound consequences for practical quantum computation and cryptography.

Stefano Pironio is in the Laboratoire d'Information Quantique, Université Libre de Bruxelles, 1050 Brussels, Belgium.
e-mail: stefano.pironio@ulb.ac.be

Trusted entanglement

DORIT AHARONOV

Schrödinger's cat is a popular image of a large quantum system. A wild tiger, however, might be more appropriate. After all, describing the quantum state of as few as 1,000 quantum spins may require $2^{1,000}$ parameters — more than the estimated number of particles in the Universe! These exponentially complex quantum states are exactly what future quantum computers will be using to achieve impressive speed-ups over classical computations. But this increase in complexity is a double-edged sword: it also means that classical systems cannot simulate complicated quantum systems in any reasonable amount of time and space, and so cannot predict their behaviour nor test whether they behave as expected⁶. And there is

good reason for not trusting quantum devices: they are extremely fragile, complex and difficult to control. Can we leash the 'quantum tiger'? Can we test whether complex quantum systems behave as they should, while trusting only our good old classical devices? Reichardt and colleagues prove that, miraculously, the answer is yes.

The authors' starting point is the CHSH game³, in which two non-communicating parties play against a referee (see Fig. 2 of the paper¹). Classical players can win only 75% of the time, but if they share a special quantum state known as the Einstein–Podolsky–Rosen (EPR) quantum state their probability of winning becomes 85%. This result is a manifestation of what Einstein called “spooky action at a distance”, also known as quantum entanglement. It provides a way of testing whether a non-communicating two-party system is in a quantum-mechanical state: play the CHSH game repeatedly, each time with the same initial state, and see whether the players win more than 75% of the games.

Now, let's reverse this logic. It turns out that if the players win 85% of the games, then their initial shared state must have been the EPR state. The main technical contribution of Reichardt *et al.* is a robust, multi-game version of this claim: if the two players play many CHSH games in sequence, starting with a shared multi-particle initial state, and win close to the optimal 85% of the games, then the entire initial state of the two players must be close to a collection of many independent EPR states. This implies much more than verifying the 'quantumness' of a system — it certifies a particular state of a large entangled quantum system, and it does so simply by posing a sequence of classical 'questions and answers' to the system being tested (Fig. 1).

Certifying entanglement of many-particle quantum systems has an important implication for high-security cryptography. Quantum-key distribution (QKD), the pinnacle of quantum cryptography, is a protocol that, remarkably, allows two parties to communicate secretly even if the entire world is trying to eavesdrop. However, realizations of this protocol are not automatically secure because of imperfections in the devices. For example, the first QKD apparatus⁸ emitted sounds that revealed information about the secrets being communicated, rendering it secure only against deaf eavesdroppers. Implementations of QKD have repeatedly been found to be insecure and to require corrections because of such issues. In 1998, Mayers and Yao envisioned⁹ using entanglement certification to achieve 'device-independent' QKD, which is secure even if the quantum devices that are used by the two parties to communicate were manufactured by the eavesdropper herself. After 15 years of important but partial progress by other researchers, Reichardt and colleagues have finally made the missing theoretical leap towards this goal: they describe a QKD protocol

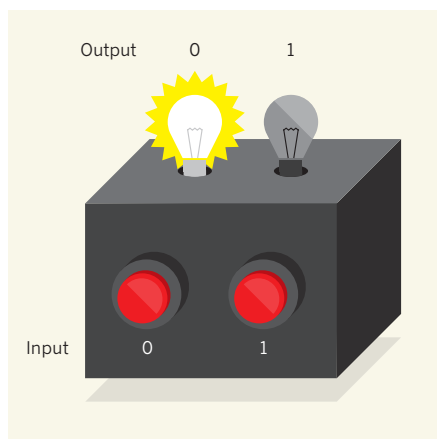


Figure 1 | Classical interaction with a quantum system. Reichardt *et al.*¹ model an arbitrarily complex quantum system as a 'black box' with simple classical inputs and outputs. An experimentalist can probe the system only by pushing button 0 or button 1, and the system outputs only two possible answers, 0 or 1, corresponding to the left or right light flashing.

and prove that it is secure even when the devices have been maliciously designed.

But there is more to it. The authors' protocol can be extended to certify the correct time-evolution of entanglement into quantum states that are considerably more complex than a collection of independent EPR states. In other words, their extended protocol certifies that a general quantum computation was performed as claimed. How can a classical experimentalist verify that such quantum states are generated even though they are much too complex for him or her to write down? This task has previously been achieved¹⁰ using a 'slightly quantum-mechanical' test. Reichardt *et al.* cleverly

provide a completely classical test, using an approach similar to that of a policewoman interrogating two thieves about a crime she knows nothing about; she looks for inconsistencies in their answers, preventing them from coordinating. The only assumptions in the authors' work are that the quantum computer being tested can be divided into two non-interacting parts, and that the tester can communicate privately with each part.

Reichardt and colleagues' protocols are yet to be made practical, that is, fault tolerant and more efficient. However, they provide a proof of principle that hands-off testing of the inner workings of arbitrarily complex quantum systems is possible. Implementing these protocols will allow new and considerably more stringent tests of quantum-information-processing devices than previously performed. ■

Dorit Aharonov is in the School of Computer Science & Engineering, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel.
e-mail: doria@cs.huji.ac.il

1. Reichardt, B. W., Unger, F. & Vazirani, U. *Nature* **496**, 456–460 (2013).
2. Bell, J. S. *Physics* **1**, 195–200 (1964).
3. Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. *Phys. Rev. Lett.* **23**, 880–884 (1969).
4. Pironio, S. *et al. Nature* **464**, 1021–1024 (2010).
5. McKague, M., Yang, T. H. & Scarani, V. *J. Phys. A* **45**, 455304 (2012).
6. Aharonov, D. & Vazirani, U. in *Computability: Turing, Gödel, Church, and Beyond* (eds Copeland, B. J., Posy, C. J. & Shagrir, O.) (MIT press, in the press).
7. Bennett, C. H. & Brassard, G. in *Proc. IEEE Int. Conf. Computers, Systems and Signal Processing*, Bangalore, 175 (IEEE, 1984).
8. Bennett, C. H., Bessette, F., Salvail, L. & Smolin, J. *J. Cryptol.* **5**, 3–28 (1992).
9. Mayers, D. & Yao, A. *Quant. Inf. Comp.* **4**, 273–286 (2004).
10. Aharonov, D., Ben-Or, M. & Eban, E. *Proc. Innov. Comp. Sci.*, Beijing, 453–469 (Tsinghua Univ. Press, 2010).

AUTOIMMUNITY

Rubbing salt in the wound

The ability of sodium chloride to induce enzymatic activity that leads to the generation of pathogenic T_H17 immune cells implicates salt as a possible factor that might exacerbate autoimmune disease. SEE LETTERS P.513 & P.518

JOHN J. O'SHEA & RUSSELL G. JONES

The role of the immune system is to protect our bodies from viral, bacterial, fungal and parasitic infections. But, sophisticated as this system is, it can go awry. One consequence is autoimmunity, a diverse collection of disorders in which the immune system turns against the host. Genetics and gender undoubtedly play key parts in the

susceptibility to autoimmune diseases, but environmental factors are also important. In this issue, Kleinewietfeld *et al.*¹ (page 518) and Wu *et al.*² (page 513) provide provocative data implicating a novel component in this mix: salt*.

The stories focus on a crucial orchestrator of immune responses — the CD4⁺, or

*This article and the papers under discussion^{1,2} were published online on 6 March 2013.

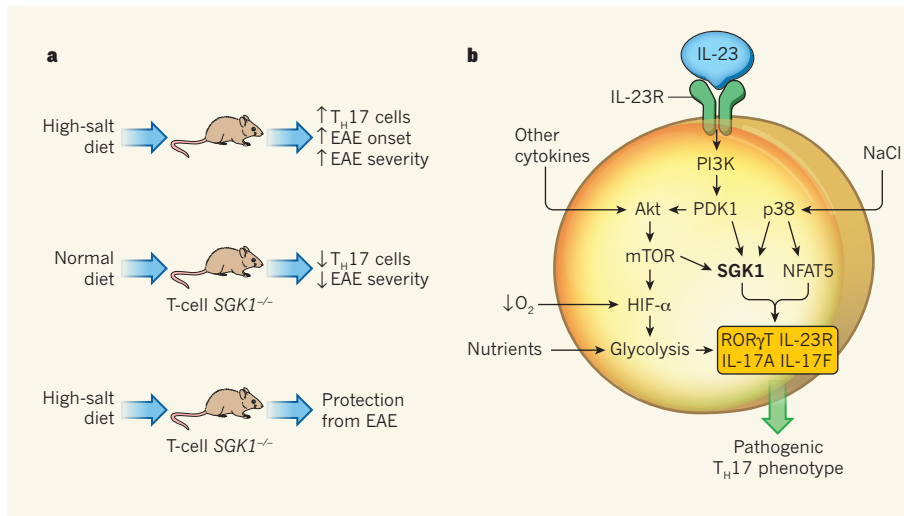


Figure 1 | SGK1 and the differentiation of T_H17 cells. **a**, Kleinewietfeld *et al.*¹ and Wu *et al.*² provide evidence that a high-salt diet can enhance the differentiation of a class of immune cell called T_H17 cells, and exacerbate disease in a mouse model of multiple sclerosis called experimental autoimmune encephalitis (EAE). They also show that mice whose T cells lack the enzyme SGK1 (T-cell SGK1^{-/-}) display reduced disease severity and are protected from NaCl-exacerbated EAE. **b**, The authors demonstrate that extracellular NaCl concentration and signalling through the IL-23 receptor both influence the activity of SGK1 to drive expression of pathogenic T_H17 -cell characteristics, which include the production of the cytokines IL-17A and IL-17F and enhanced expression of the IL-23 receptor (IL-23R) and the transcription factor ROR γ T (encoded by *Rorc*). However, this finding must be considered in the context of other environmental factors, such as oxygen and nutrient provision. These influence signalling pathways and glycolytic metabolism in ways that regulate not only T_H17 -cell differentiation, but also that of other classes of T cell.

'helper', T cells. These cells regulate immune responses through their ability to differentiate into distinct classes of cell according to the nature of the offending pathogen. In the past decade, increasing attention has focused on a subset of CD4⁺ T cells, commonly known as T helper 17 (T_H17) cells³⁻⁶, which secrete molecules belonging to the IL-17 group of cell-signalling compounds called cytokines. Cells that produce IL-17 are prominent in the gut, where they influence its barrier function and help to protect against extracellular pathogens and fungi. However, these helpers can also be traitors — T_H17 cells are important drivers of autoimmune disease, and have inflammatory properties.

The present studies tell their stories in different ways, but both show that an elevated sodium chloride (NaCl) concentration (40–80 millimolar) in an otherwise isotonic culture medium promotes the differentiation of CD4⁺ T cells into T_H17 cells *in vitro*. Perhaps the most provocative experiments relate to an *in vivo* correlate of this finding. The authors demonstrate that a high-salt diet accelerates neuropathology in experimental autoimmune encephalomyelitis (EAE), a mouse model of the autoimmune disease multiple sclerosis. They used inhibitors, interfering RNA molecules and knockout mice to test the role of cellular signalling pathways in these processes, and link the regulation of NaCl and T_H17 differentiation with the transcription factor NFAT5 and the protein-kinase enzymes p38 and SGK1 (Fig. 1).

This makes sense, because p38 is an evolutionarily conserved kinase that is activated by changes in cellular osmolarity, and NFAT5 and SGK1 are both substrates of p38.

The authors also find that SGK1 is expressed in T_H17 cells and is induced by NaCl, and show that mice lacking this kinase in their T cells have impaired expression of IL-17-family cytokines and of a receptor for another cytokine molecule, IL-23. When they tested these knockout mice in the EAE model, they found that the lack of SGK1 also leads to reduced neuropathology. SGK1 has been implicated in inflammatory pathways before: it is known to inactivate the transcription factor Foxo1. Accordingly, Foxo1-deficient T cells have higher levels of IL-17 and IL-23-receptor expression.

In considering these studies, it is appropriate to re-emphasize that IL-17 and T_H17 cells are not always the villains; they also protect us from a universe of true villains. In the same vein, it should also be pointed out that not all T_H17 cells are alike: although some IL-17-producing T cells mediate immune pathology, others do not. IL-23 is a key cytokine in generating pathogenic T_H17 cells⁷, and the authors of both papers note that NaCl and SGK1 seem to contribute to the generation of T_H17 cells that have pathogenic potential. However, many cells other than CD4⁺ T cells, including innate immune cells and γ/δ T cells, also produce IL-17 and related cytokines⁸. Although a high-salt diet may indeed worsen autoimmune disease,

the data provided do not establish exactly which cells NaCl works on to achieve this.

An additional point is that autoimmune diseases are heterogeneous, and the benefit achieved by blocking IL-17 is variable. Inhibiting IL-17 is useful in treating psoriasis, but less so in inflammatory bowel disease; the jury is still out on whether targeting IL-17 will be of help in treating multiple sclerosis. It is also worth noting that the studies by Kleinewietfeld *et al.*¹ and Wu *et al.*² show that NaCl exacerbates an artificially created disease; there are no data indicating that dietary salt promotes or worsens spontaneous disease.

The complex interaction of the factors that regulate helper-T-cell differentiation must also be taken into account when considering these results (Fig. 1). SGK1 is a member of the AGC family of protein kinases, and is homologous to the enzyme Akt⁹. Akt has well-documented effects on cell survival, metabolism and helper-T-cell differentiation. Moreover, Akt and SGK1 share upstream activators, including the enzymes PI3K and PDK1, and downstream substrates. Key molecules, such as mTor and Foxo1, are also influenced by diverse factors and have complex effects on T-cell function¹⁰. Furthermore, helper T cells are influenced by nutrient availability and by the oxygen-sensitive transcription factor HIF-1 α , which suggests a close link between metabolism and differentiation. Similarly, p38 is also a recognized regulator of helper T cells¹¹. This is pertinent to the present studies, because the authors' results suggest that the functions of SGK1 and NaCl are not entirely congruent: SGK1 positively regulates IL-17, but negatively regulates the genes *Ifng*, *Tbx21*, *Il4*, *Il13*, *Gata3*, *Il2* and *Il9*, whereas NaCl positively regulates IL-17, *Ifng*, *Tbx21*, *Il2* and *Il9*.

Thus, dietary salt is just one of many factors that influence helper T cells; cytokines, the microbiota, diet, metabolism and other diverse environmental factors are all important too^{10,12,13}. The bottom line is that these kinases and transcription factors represent key nodes for many receptors and signalling pathways that integrate a vast array of stimuli. So, although these are exciting and provocative data, it is clearly premature — as also pointed out by both sets of authors — to state that dietary salt influences autoimmune disease in humans and that this is mediated by T-cell-induced production of IL-17. However, the work should spur investigation of tangible links between diet and autoimmune disease in people. In doing so, it will be essential to conduct formal, controlled clinical trials. Fortunately, the risks of limiting dietary salt intake are not great, so it is likely that several such trials will be starting soon. ■

John J. O'Shea is in the Intramural Research Program, National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda, Maryland 20892-1616, USA.

Russell G. Jones is at the Rosalind and Morris Goodman Cancer Research Centre, Department of Physiology, McGill University, Montreal, Quebec H3G 1Y6, Canada. e-mails: osheaj@arb.niams.nih.gov; russell.jones@mcgill.ca

1. Kleinewietfeld, M. et al. *Nature* **496**, 518–522 (2013).
2. Wu, C. et al. *Nature* **496**, 513–517 (2013).
3. Weaver, C. T., Hatton, R. D., Mangan, P. R. & Harrington, L. E. *Annu. Rev. Immunol.* **25**, 821–852 (2007).
4. Stockinger, B., Veldhoen, M. & Martin, B. *Semin. Immunol.* **19**, 353–361 (2007).
5. Kolls, J. K. & Linden, A. *Immunity* **21**, 467–476 (2004).

6. Miossec, P. & Kolls, J. K. *Nature Rev. Drug Discov.* **11**, 763–776 (2012).
7. Cua, D. J. et al. *Nature* **421**, 744–748 (2003).
8. Sutton, C. E., Mielke, L. A. & Mills, K. H. G. *Eur. J. Immunol.* **42**, 2221–2231 (2012).
9. Bruhn, M. A., Pearson, R. B., Hannan, R. D. & Sheppard, K. E. *Growth Factors* **28**, 394–408 (2010).
10. Honda, K. & Littman, D. R. *Annu. Rev. Immunol.* **30**, 759–795 (2012).
11. Dong, C., Davis, R. J. & Flavell, R. A. *Annu. Rev. Immunol.* **20**, 55–72 (2002).
12. Wang, R. & Green, D. R. *Nature Immunol.* **13**, 907–915 (2012).
13. Stockinger, B., Hirota, K., Duarte, J. & Veldhoen, M. *Semin. Immunol.* **23**, 99–105 (2011).

PARTICLE PHYSICS

Minimalism triumphant

The discovery of a particle that looked like the Higgs boson marked a milestone for physics. Results reported since then are strikingly consistent with expectations for the Higgs particle of the minimal standard model of particle physics.

FRANK WILCZEK

Since the announcement last July that a new kind of particle had been discovered at the Large Hadron Collider (LHC) at CERN, Europe's particle-physics laboratory near Geneva in Switzerland, a much fuller portrait of that particle has emerged. The two main experimental collaborations, ATLAS and CMS, reported a host of measurements in papers and presentations at last month's Moriond conference in La Thuile, Italy^{1,2}. So far, all results remain consistent with the interpretation that the new particle is the Higgs boson anticipated in the minimal implementation of electroweak symmetry breaking in the standard model of particle physics.

The Higgs particle is a rare and fleeting physical phenomenon. Even at the LHC, the particle is produced in less than one-billionth of the proton–proton collisions, and it is highly unstable — its lifetime is inferred to be about 10^{-22} seconds. To appreciate the significance of the Higgs particle, it is necessary to put it in its proper context — the Higgs mechanism.

A central assumption of the standard model, inferred from many experiments, is that the basic forces of nature — the strong, weak and electromagnetic forces, as well as gravity — are mediated by quantum fields of spin 1 or (for gravity) spin 2. It is challenging to accommodate that assumption theoretically, in consistent equations. Naive attempts founder because they predict the existence of violent (quantum) fluctuations in the fields at short distances, which lead to a plague of mathematical infinities in calculations of physical quantities. These difficulties can be avoided only in theories in which the fields have

enormous symmetry, called gauge symmetry.

Gauge symmetry, however, seems to require that the most basic manifestations of the gauge fields, the minimal concentrations of energy or quanta of the fields, are particles with zero mass. That property holds true in many cases: the photons of electromagnetism, the colour gluons of the strong interaction and the graviton of gravity do seem to have zero mass. But W and Z bosons, the quanta of the fields that are responsible for the weak interaction, have substantial masses.

The Higgs mechanism provides a way out of this difficulty. The key observation is that gauge symmetry requires the W and Z bosons to have zero mass only in empty space. Material can slow them down, screen their influence and make them behave as if they have non-zero mass. If an appropriate material fills all space uniformly and stably, the W and Z bosons will never escape its influence — and they will always be observed to have non-zero mass. The hypothesis that such a material does, in fact, fill space is the essence of the Higgs mechanism. But does this material exist? And, if so, what is it made out of?

The triumph of the standard-model account of weak interactions, which relies on the Higgs mechanism, has long provided overwhelming, if circumstantial, evidence that the material exists. In recent months, we have learned what it is made out of. Among all the logical possibilities for the new material, the simplest and most economical proposal defines the 'minimal standard model'. In this model, the cosmic material is made from just one ingredient. The terminology in this subject is both confused and in flux, but here the term Higgs particle is used to refer to the unique particle that is introduced to



50 Years Ago

'Internal circulations within liquid drops' — It has long been recognized that under certain conditions some sort of axisymmetric flow is induced within liquid drops as they pass through a viscous medium ... Although an internal circulation theory has been very attractive in considerations of meteorological phenomena, there appears to be little experimental evidence to substantiate such a theory ... In this communication we discuss a technique which affords velocity measurements and at the same time outlines the vortical core ... The streamlines within the drop are recorded photographically by means of dye trails. The figure indicates the streak-lines due to the motion within water drops moving through mineral oil ... We have obtained a great deal of velocity data with this method.



From *Nature* 27 April 1963

100 Years Ago

The twinkling of stars may be imitated in the dark-room. If a small light be looked at in a dark-room, as, for instance, that coming through the smallest diaphragm of my colour perception lantern, ... care being taken not to move the eye, the light will appear to twinkle like a star. It will be noticed that pale bluish-violet circles start at the periphery of the field of vision, and, gradually contracting, reach the centre. On reaching the centre the light brightens. If the circles stop the light disappears. The colour of the circle is the same for white light or any colour.

From *Nature* 24 April 1913

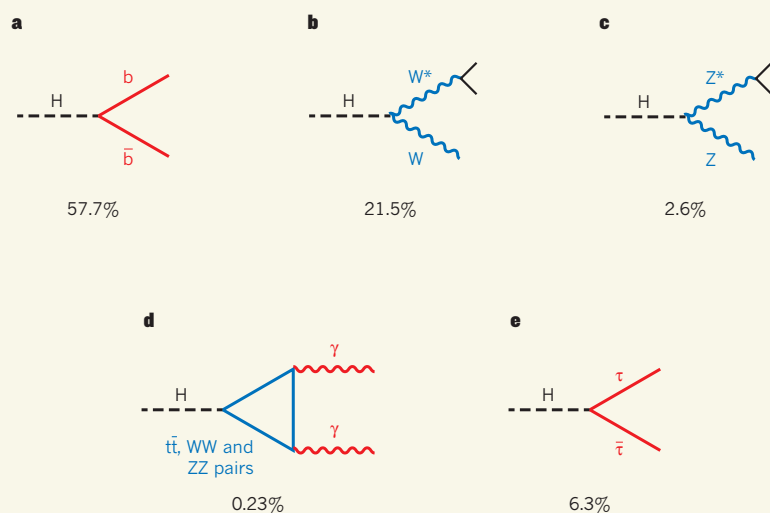


Figure 1 | Decay modes of the Higgs particle. The numbers represent the percentage probabilities for each decay mode, as calculated in the minimal standard model. **a**, Bottom quark-antiquark ($b\bar{b}$) particle pairs. **b**, W boson pairs. The H is not heavy enough to decay into two W bosons, so one (W^*) never materializes as such, but ‘decays’ almost before it is actually produced. The other decays normally. **c**, Z boson pairs, conceptually similar to W boson pairs. **d**, Photon ($\gamma\gamma$) pairs. Because photons do not couple directly to H, this decay proceeds by an indirect mechanism (denoted by the triangle) that can involve a top quark-antiquark ($t\bar{t}$) pair and W and Z boson pairs. **e**, Tau lepton-antilepton ($\tau^+\tau^-$) pairs. Other decay channels are possible, notably gluon pairs and charm quark-antiquark pairs, but are more challenging to access experimentally, because energetic gluons and charm quarks are easily produced by other means, raising severe signal-to-noise issues.

complete the minimal standard model.

We can infer a great deal about how the Higgs particle interacts with other forms of matter. After all, because we are embedded in a cosmic material made from Higgs particles, we have observed their en masse effects on matter for a long time. In fact, all properties of the Higgs particle — including its spin and its rate of production — can be, and were, predicted given only its mass.

The Higgs particle can be produced in several ways, and it can decay in several ways (Fig. 1). This wealth of possibilities affords many opportunities for observations to test the underlying theory. If the Higgs particle were produced in isolation and in a clean environment, it would be straightforward to observe the rates of each production mechanism and each decay mode, and thereby to test the theory in full detail. High-energy proton-proton collisions, however, are far from that ideal. Even in the rare Higgs-containing collisions, very few of the dozens of particles that are produced have anything to do with the Higgs particle. Tremendous effort has gone into understanding these ‘backgrounds’, which themselves reflect fundamental processes. As Richard Feynman once said, “yesterday’s sensation is today’s calibration”, but in these extraordinary conditions successful anticipation of what happens 99.9999999% of the time is as remarkable as the new information contained in the remaining 0.0000001%.

Enumeration of all the combinations of production and decay processes that have been observed by the ATLAS and CMS collaborations is not appropriate here. I will briefly describe the two most mature cases, and mention even more briefly a few other results.

The announcements of the initial discovery at CERN were based mainly on observation of an excess signal in the two-photon ($\gamma\gamma$) decay channel (Fig. 1d) at effective masses of about 125 billion electronvolts, relative both to computed backgrounds and to the measured background at nearby mass values. Although the $\gamma\gamma$ decay mode is rare for the Higgs particle, it is also difficult to produce high-mass photon pairs by other means, so the background is suppressed. Furthermore, it is possible to measure the energy and direction of photons quite accurately, which has enabled rapid progress in the study of this decay channel. So far, all results^{1,2} are consistent with expectations for a Higgs particle having a mass of 125 billion electronvolts. Initial hints that the rate of $\gamma\gamma$ production through Higgs-particle decay, relative to backgrounds, might exceed theoretical expectations have softened.

Although it too represents a small decay fraction, the ZZ decay mode (Fig. 1c) is particularly favourable for study. This is because the Z boson often decays into two charged leptons (electron or muon pairs), whose energy and momentum can be measured

accurately. This allows full reconstruction of the underlying process, and comparison of energy and momentum distributions with theoretical predictions. These extra handles help us to address another fundamental property of the Higgs particle: its spin. Theory predicts that the particle should have the quantum numbers of ‘empty’ space, namely spin 0 and positive parity, because it is a quantum of (apparently) empty space. Detailed study of angular and energy distributions in the ZZ decay have provided strong evidence in favour of spin 0 and positive parity, as anticipated.

In the $\gamma\gamma$ and ZZ channels, quantitative comparison of theory and experiment is at the level of a few tens of per cent. The WW and tau lepton-antilepton ($\tau^+\tau^-$) decay channels (Fig. 1b,e) have also been observed, although in the case of the latter the precision is less good.

Two dominant themes emerge from these findings. The first is that the LHC machine works beautifully, and that the experimental groups are exploiting it brilliantly. Over the course of a few months, tentative sighting of the Higgs particle has matured into its multi-featured, quantitative portrait. The second is that, so far, every aspect of the emerging portrait is consistent with expectations for the Higgs particle of the minimal standard model.

The new results challenge several widely mooted speculations. Models that postulate several main ingredients to the cosmic material (for example, multi-Higgs models), or that postulate complex dynamics to explain the W and Z boson masses (such as Technicolor, extra-dimension and brane-world models) seem less credible, as simplicity and minimalism carry the day.

Concerning the most popular and, in my opinion, most promising speculation about physics beyond the minimal standard model that might be accessed at the LHC, supersymmetry, the message is mixed. The observed mass of the Higgs particle is quite low, relative to a priori expectations, as supersymmetry requires. But implementations of supersymmetry that allow a mass as large as 125 billion electronvolts seem to require quite heavy masses for the supersymmetric partner particles, squarks and leptons, perhaps 10–100 teraelectronvolts. Unfortunately, but ‘conveniently’, that puts them beyond the reach of the LHC. (Other superpartners, the gauginos, might be accessible.)

Focus point³ or split⁴ supersymmetry models, which anticipated this possibility, also suppress many other logically possible, but unobserved, signatures of supersymmetry. The flip side of those negative virtues is that these models compromise one widely advertised advantage of supersymmetry, its potential to ease the hierarchy problem. (The hierarchy problem is the ‘unnaturally’ tiny value of the W boson mass, relative to the fundamental Planck mass.) With that gone, only unification of

couplings⁵ — an explanation of the relative powers of the strong, weak, electromagnetic and gravitational forces — remains as a firm, reasonably quantitative motivation for supersymmetry.

We are unlikely to see notable, qualitatively new results from the LHC in the immediate future because the machine will be out of commission for at least a year while it is upgraded to allow higher collisional energy and luminosity. The second-generation LHC will empower greater accuracy in all the checks of minimalism, and possibly finally deliver supersymmetry — or an unanticipated surprise! ■

Frank Wilczek is in the Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
e-mail: wilczek@mit.edu

1. Atlas Experiment — Public Results. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/HiggsPublicResults>
2. CMS Higgs Physics Results. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsHIG>
3. Feng, J., Matchev, K. & Moroi, T. *Phys. Rev. D* **61**, 075005 (2000).
4. Arkani-Hamed, N. & Dimopoulos, S. *J. High Energy Phys.* JHEP06(2005)073 (2005).
5. Dimopoulos, S., Raby, S. & Wilczek, F. *Phys. Rev. D* **24**, 1681–1683 (1981).

HIV

Roadmaps to a vaccine

More than 30 years since the AIDS pandemic began, there is still no effective vaccine. But analysis of broadly acting, potent human antibodies obtained from single cells suggests a rational approach to vaccine development. SEE ARTICLE P.469

HUGO MOUQUET & MICHEL C. NUSSENZWEIG

Our understanding of how humans respond to HIV has been revolutionized by the introduction of techniques for isolating anti-viral antibodies from single cells¹. Such methods have led to the discovery of naturally occurring, potent antibodies that can neutralize a broad range of HIV viruses, and prevent² and suppress³ infection. These findings, combined with the association between antibody responses and protection from infection that was identified in a human trial⁴ of the vaccine RV144, have reinvigorated the quest for antibody-based HIV vaccines. However, it has also become clear that anti-HIV antibodies undergo unusually high levels of mutation^{1,5}, which represents a potential stumbling block for vaccine development. Among four recent studies^{6–9} that address this subject is a paper in this issue, in which Liao *et al.*⁹ (page 469) track antibody and viral evolution during one patient's response to HIV*.

Antibodies are produced by the B lymphocytes of the immune system. The receptors on the surface of each circulating B cell are unique, enabling an immune response to any foreign structure. When a B cell encounters an entity that matches its receptor, it is stimulated to proliferate and secrete antibodies against that structure. Although B-cell genes frequently undergo somatic (non-germline) mutation to increase the affinity of the antibodies they

produce, anti-HIV antibodies are unusual in that they are highly somatically mutated — they are therefore quite different from those encoded by the B cells that initially respond to the infection^{1,10}. Furthermore, these mutations seem to be required for the antibodies to bind to heterologous viral-envelope proteins (those expressed on most HIV viruses)^{5,10}. If B cells that express the germline antibody precursor do not bind to the antigen, how are they stimulated in the first place, and why do the antibodies need so many mutations? Answering these questions is of fundamental importance in attempts to reproduce this antibody-development process by vaccination.

Some patients with HIV develop broadly neutralizing antibody activity, but only 2–4 years after infection. Scrutiny of the antibodies produced by single human B cells¹ showed that these broadly neutralizing responses are due to a combination of antibodies in some individuals, and to single, potent antibodies in others². In an attempt to dissect the natural pathways that lead to the generation of these antibodies, Liao *et al.* studied a patient who developed broad and potent antibodies.

The authors investigated the co-evolution of the HIV-1 virus and the broadly neutralizing antibodies for 34 months from the start of the infection. They isolated a virus-specific antibody named CH103, and clonal variants of it, from single memory B cells that were obtained using a fluorescently tagged viral-envelope protein as bait^{3,5,11}. CH103 neutralizes 55% of HIV-1 isolates and targets the site

on the virus that binds to CD4 molecules on the surface of T cells (the immune cells that HIV infects). Like other antibodies in this class^{5,11}, CH103 is highly somatically mutated, and its unmutated germline precursor fails to bind to heterologous HIV-1 envelope proteins⁹.

One of Liao and colleagues' key findings is that the germline precursor antibody of CH103 has high affinity for the envelope protein expressed by the founder virus that infected the individual. The authors suggest that a progenitor B cell that expresses this germline antibody might only be stimulated to respond if it is presented with the envelope proteins of the founder virus, or similar proteins. The idea that certain envelope proteins are more likely to induce broadly neutralizing antibodies is supported by experiments in macaques showing that specific envelopes induce such responses to simian HIV, whereas others do not¹².

However, simply initiating the antibody response is not sufficient for effective immune defence. It takes time and unusually large numbers of somatic mutations for antibody breadth and potency to develop. Liao *et al.* reconstructed the CH103 clonal lineage by using samples that went back to the time of infection. Although all members of the lineage recognized and neutralized the founder virus, the affinity and neutralizing activity against heterologous viruses gradually increased through the accumulation of somatic mutations. The authors also found that, as previously described for glycan-dependent broadly

“These data suggest a molecular explanation for why broadly neutralizing anti-HIV antibodies take 2–4 years to develop.”

neutralizing antibodies¹³, viral diversification and the emergence of ‘escape mutants’ (those with mutations in the site targeted by the antibody) preceded the development of antibody breadth. By studying a crystal structure of the CH103 antibody in complex with its envelope protein target, Liao *et al.* showed that HIV escapes antibody pressure by mutating amino-acid residues in and around the CD4 binding site. These resistant viruses then elicit further somatic mutation and ‘affinity maturation’ of CH103 antibody variants, resulting in greater neutralization breadth of the antibody response.

The reason for the high level of somatic mutation required to produce broadly acting, potent anti-HIV antibodies has recently been investigated⁶. Under normal circumstances, high affinity of an antibody for its target is usually achieved after the accumulation of 10–15 mutations in the complementarity-determining region of the antibody that forms the antigen contact site. However, broad and

*This article and the paper under discussion⁹ were published online on 3 April 2013.

potent anti-HIV antibodies contain 40–100 somatic mutations^{1,5,6,11} that span both the complementarity-determining region and the relatively constant, and mutation-resistant, framework regions. Experiments in which mutations in the framework regions were selectively reverted showed that these mutations are necessary for the evolution of broad and potent anti-HIV antibodies⁶. These structural alterations in the antibody were found to contribute to direct contacts with the virus and to enhanced flexibility of the antibody structure, both of which are required for optimal breadth and potency.

Combined with Liao and colleagues' findings, these data suggest a molecular explanation for why broadly neutralizing anti-HIV antibodies take 2–4 years to develop. Moreover, they indicate that an effective vaccine may require shepherding of B-cell responses

through multiple rounds of the natural antibody maturation and mutation process, using naturally derived viral envelopes that induce the production of broad and potent antibodies in people with HIV. A recently suggested^{7,8} alternative, non-mutually exclusive approach is to design specific 'immunogen' molecules that would bind to and activate B cells that produce the germline precursors of broadly neutralizing antibodies. Whether such roadmaps can be used to design effective vaccine strategies has yet to be determined, but they present a strong and testable route to addressing the main challenges of creating an antibody-based HIV-1 vaccine. ■

Hugo Mouquet and Michel C. Nussenzweig are at the Rockefeller University and the Howard Hughes Medical Institute, New York, New York 10065, USA.

e-mails: hmouquet@rockefeller.edu; nussen@rockefeller.edu

1. Corti, D. & Lanzavecchia, A. *Annu. Rev. Immunol.* **31**, 705–742 (2013).
2. Scheid, J. F. *et al. Nature* **458**, 636–640 (2009).
3. Klein, F. *et al. Nature* **492**, 118–122 (2012).
4. Haynes, B. F. *et al. N. Engl. J. Med.* **366**, 1275–1286 (2012).
5. Scheid, J. F. *et al. Science* **333**, 1633–1637 (2011).
6. Klein, F. *et al. Cell* <http://dx.doi.org/10.1016/j.cell.2013.03.018> (2013).
7. McGuire, A. T. *et al. J. Exp. Med.* <http://jem.rupress.org/content/early/2013/03/19/jem.20122824> (2013).
8. Jardine, J. *et al. Science* <http://dx.doi.org/10.1126/science.1234150> (2013).
9. Liao, H. X. *et al. Nature* **496**, 469–476 (2013).
10. Mouquet, H. *et al. Nature* **467**, 591–595 (2010).
11. Wu, X. *et al. Science* **329**, 856–861 (2010).
12. Shingai, M. *et al. Proc. Natl Acad. Sci. USA* **109**, 19769–19774 (2012).
13. Moore, P. L. *et al. Nature Med.* **18**, 1688–1692 (2012).

created by the chains of magnetic nanocrystals in the cells. Le Sage and colleagues' experimental set-up allowed them to simultaneously acquire magnetic maps and optical images of the bacteria. In this way, they could compare the recorded magnetic fields with the positions of the cells, map the positions of the chains of magnetic nanoparticles (see Fig. 4 of the paper¹) and quantify the magnetic moments of the chains.

The importance of the technique for studying biomagnetic structures lies in the fact that both magnetic and optical images can be collected with a spatial resolution of about 400 nanometres from a population of cells across a wide field of view — spanning $100\mu\text{m} \times 30\mu\text{m}$. Although other approaches provide better spatial resolution for imaging magnetic fields in bacteria^{3,5,10}, at present these methods cannot be used under ambient conditions and for imaging multiple cells across such a large field of view in real time. Le Sage and colleagues' study opens up the possibility of dynamic imaging of the development of magnetic fields in bacteria as their chains of magnetic crystals grow.

Another potential application would be to screen non-magnetic mutant bacteria produced in genetic-engineering studies aimed at understanding the biological mechanisms that control the growth of magnetic nanocrystals inside cells¹¹. The sharing of magnetic nanoparticles between daughter cells during cell division could also be studied. In addition to understanding magnetic nanocrystal formation by bacteria, it may be possible to use the method to reveal the presence and evolution of putative magnetic structures in the tissues of more complex organisms, including insects, birds and humans, under ambient conditions.

Some words of caution are warranted before making excessively bold predictions about

IMAGING

Magnetic bacteria on a diamond plate

A new approach has been used to image magnetic fields in living cells of magnetotactic bacteria. The technique could be applied to study the dynamics of magnetism in other biological systems. SEE LETTER P.486

MIHÁLY PÓSFAI &
RAFAL E. DUNIN-BORKOWSKI

Just as schoolchildren sprinkle iron filings on a sheet of paper placed over a magnet to visualize the magnetic field around the magnet, scientists who are interested in magnetism strive to image the magnetic fields within and around objects across a wide range of spatial and temporal scales. Although many different magnetic imaging techniques are now available, imaging micro- and nanoscale magnetic fields in living organisms is still challenging. On page 486 of this issue, Le Sage *et al.*¹ describe an advanced optical magnetic imaging technique which they use to study the three-dimensional magnetic fields that originate from chains of magnetic nanocrystals inside the living cells of magnetotactic bacteria.

Many organisms contain magnetic nanocrystals inside their bodies; some use them to navigate in magnetic fields, whereas others use them to harden or protect their tissues. Magnetotactic bacteria are the simplest organisms that are known to contain magnetic nanocrystals. Their delicate internal chains of tailor-made iron oxide or iron sulphide particles have attracted intense scientific interest since their discovery², and are often used as

nanoscale natural laboratories to develop and test magnetic imaging techniques^{3–6}.

The fundamental principles of the technique that Le Sage *et al.* use have been known for some time⁷ and have been applied to map magnetic-field variations on the nanoscale^{8,9}.

“The study opens up the possibility of dynamic imaging of the development of magnetic fields in bacteria as their chains of magnetic crystals grow.”

They involve detecting changes in the quantum spin states of crystallographic defects called nitrogen–vacancy centres in a diamond chip (a nitrogen atom and a vacancy substitute for two neighbouring carbon atoms in the diamond crystal lattice). The novelty of the authors' study lies in using this approach to image magnetic fields in living microorganisms.

When the authors placed magnetotactic bacteria on a diamond surface, they found that the cells' magnetic fields affected characteristic signals, known as electron spin resonance frequencies, of the nitrogen–vacancy centres in the diamond. They detected such signals using an optical beam, and reconstructed all vector components of the magnetic field

future uses of this imaging approach. First, the spatial resolution depends on the distance between the diamond surface and the source of the magnetic field. Submicrometre resolution in the recorded magnetic images was achieved only when the cells were dried (and necessarily dead) on the diamond surface. By contrast, when the bacteria were alive in a liquid environment, the cells were farther away from the nitrogen-vacancy centres and the resolution deteriorated.

Second, it may be possible to adapt several other magnetic imaging techniques, including SQUID microscopy, electron holography and magnetic resonance imaging, to achieve similar results in ambient conditions. These methods might provide higher-spatial-resolution alternatives to the optical magnetic imaging technique used by Le Sage and colleagues. Nevertheless, at the moment, the diamond-chip-based, optical magnetic imaging approach described by the authors is the only game in town that can be used to obtain quantitative, three-dimensional nanoscale information about magnetic fields originating from living microorganisms across a large field of view. ■

Mihály Pósfai is in the Department of Earth and Environmental Sciences, University of Pannonia, 8200 Veszprém, Hungary.

Rafal E. Dunin-Borkowski is in the Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons and the Peter Grünberg Institute, Forschungszentrum Jülich, 52425 Jülich, Germany.

e-mails: mihaly.posfai@gmail.com;
rafaldb@gmail.com

1. Le Sage, D. *et al. Nature* **496**, 486–489 (2013).
2. Blakemore, R. *Science* **190**, 377–379 (1975).
3. Dunin-Borkowski, R. E. *et al. Science* **282**, 1868–1870 (1998).
4. Lam, K. P. *et al. Chem. Geol.* **270**, 110–116 (2010).
5. Proksch, R. *et al. Appl. Phys. Lett.* **66**, 2582–2584 (1995).
6. Staniland, S., Ward, B., Harrison, A., van der Laan, G. & Telling, N. *Proc. Natl Acad. Sci. USA* **104**, 19524–19528 (2007).
7. Gruber, A. *et al. Science* **276**, 2012–2014 (1997).
8. Balasubramanian, G. *et al. Nature* **455**, 648–651 (2008).
9. Kolkowitz, S., Unterreithmeier, Q. P., Bennett, S. D. & Lukin, M. D. *Phys. Rev. Lett.* **109**, 137601 (2012).
10. Degen, C. L., Poggio, M., Mamin, H. J., Rettner, C. T. & Rugar, D. *Proc. Natl Acad. Sci. USA* **106**, 1313–1317 (2009).
11. Murat, D., Quinlan, A., Vali, H. & Komeili, A. *Proc. Natl Acad. Sci. USA* **107**, 5593–5598 (2010).

GENOMICS

Zebrafish earns its stripes

The reported sequence of the zebrafish genome, together with the production of mutant strains representing more than one-third of all its protein-coding regions, will accelerate the characterization of human genes. SEE LETTERS P.494 & P.498

ALEXANDER F. SCHIER

Thousands of genes and gene variants are thought to contribute to human development, physiology and disease, but the functions of most of them are unknown. In the past 20 years, the zebrafish has emerged as a model system to investigate the function of human genes. Two papers in this issue^{1,2}, reporting the sequence of the zebrafish genome and the isolation of disruptive mutations in more than 10,000 protein-coding genes, add to other recent studies^{3–7} in providing a strong boost to this effort*.

A common approach to studying gene function is to determine how a mutation changes an organism's phenotype, which includes its anatomy, physiology and behaviour. Zebrafish embryos and larvae are ideally suited for such studies: their small size, accessibility and transparency allow analysis of thousands of live

animals at single-cell resolution. Most gene functions in zebrafish have been uncovered by 'forward genetics' approaches, in which genomic changes are induced randomly and resultant changes to phenotype are identified in later generations^{8,9} (Fig. 1a). Identifying causative mutations using this strategy is laborious, but the approach has helped to uncover genetic pathways that control processes ranging from embryonic development to heart physiology. Many of these pathways are conserved in humans, which strengthens the use of zebrafish as a model system to study human gene function.

The high-quality zebrafish genome sequence reported by Howe *et al.*¹ (page 498) greatly facilitates the identification of mutations, because it makes possible a direct comparison of mutated and normal sequences. The genome sequence also reveals that more than 75% of human genes implicated in disease have counterparts in zebrafish, providing an opportunity to analyse their roles in this model system.

*This article and the papers under discussion^{1,2} were published online on 17 April 2013.

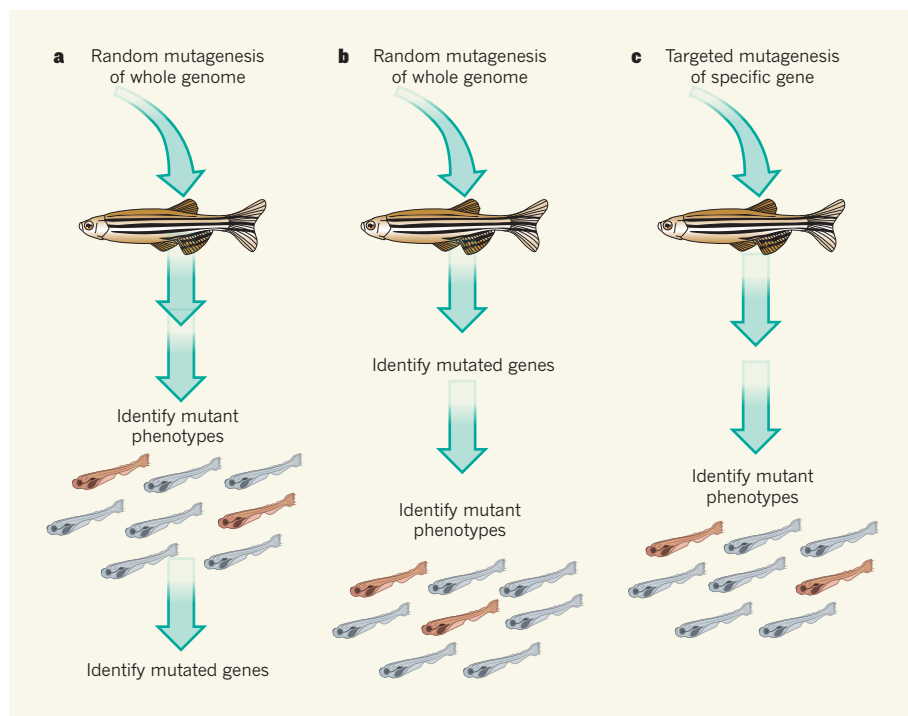


Figure 1 | Linking phenotypes to genes. Three approaches to identifying the role of specific genes in determining an organism's phenotype are commonly used in zebrafish. **a**, Forward genetics involves introducing random mutations into adults, identifying offspring with phenotypic changes and then analysing their genomes for mutations. **b**, An alternative method is to perform random mutagenesis but then to use whole-genome sequence comparisons to identify mutations in offspring before seeking phenotypic changes. **c**, In targeted approaches, mutations are introduced into specific genes and the result is then monitored in the offspring.

But how can the function of a specific human gene be studied in zebrafish? Mutations need to be introduced into the zebrafish gene counterpart; this is then followed by phenotypic analysis (Fig. 1b). Kettleborough *et al.*² (page 494) demonstrate how this can be done on a large scale. The authors subjected adult male zebrafish to random mutagenesis, and then sequenced protein-coding regions in the offspring's DNA. In 1,673 fish they identified disruptions in 10,043 genes — more than one-third of all zebrafish protein-coding genes. These mutant strains now provide a resource for the systematic analysis of the function of these genes.

An alternative approach to interfering with gene function is to focus on specific genes rather than introducing random mutations into the entire genome (Fig. 1c). The development of systems to cleave DNA at specific sites has the potential to revolutionize the way we carry out such targeted mutation. The latest approach is the CRISPR-Cas9 system, which has already been applied successfully to zebrafish^{4,5}. In this method, an RNA molecule whose sequence is complementary to part of a gene of interest guides an endonuclease enzyme to a specific DNA site, resulting in cleavage, improper repair and, therefore, mutation of the targeted gene. This system is cheap and rapid and, in contrast to large-scale mutagenesis screens, can be used by small laboratories. So it is only

a matter of time until mutated versions of most zebrafish genes are generated.

Despite these breakthroughs, we still have little idea of how many genes, when disrupted, will produce phenotypic changes. Previous forward genetic screens have estimated^{9,10} that disruption of less than 10% of zebrafish genes causes abnormalities during the first five days of embryonic and larval development. Consistent with these observations, Kettleborough *et al.*² found that only around 5% of more than 800 tested genes were required for normal development during that time period.

Does this mean that more than 90% of zebrafish genes are functionally irrelevant? Several considerations make this possibility very unlikely. First, Kettleborough and colleagues' phenotypic analysis included only obvious anatomical features, and more subtle phenotypes would have been missed. Second, many genes are required only during later stages of development or in adults. Third, at this early stage of development, defective gene function may often be masked by the contribution of gene products provided by the mother, which are present in both the egg and the embryo. Finally, genes with similar sequences often have overlapping or partially redundant functions, resulting in no or subtle defects on disruption of a single gene.

Howe *et al.* found that one-quarter of zebrafish genes have sister genes with high

sequence similarity, suggesting that this last possibility might be particularly relevant to zebrafish. However, all of these possibilities can be addressed by the sophisticated techniques now available for use in zebrafish: phenotypes can be analysed by high-resolution imaging and gene-expression profiling; mutant fish can be studied at later stages of development; and it is feasible to generate mutants that lack both maternal and embryonic gene functions or to disrupt two or three related genes in the same animal.

How will these tools and resources accelerate the study of human disease genes? The recipe seems clear: find or engineer mutations in the zebrafish counterpart of a human disease gene, analyse any phenotypic abnormalities and use high-throughput drug-screening platforms to discover or characterize small molecules that can modulate those phenotypic changes¹¹. In addition, the zebrafish genome sequence might open the door to studying old problems using new approaches. For instance, zebrafish have very high rates of genetic variation — Howe *et al.* found that 1 in 200 bases differ between strains or even between individuals. Studies in other organisms have shown that such differences can have phenotypic effects. It is therefore conceivable that zebrafish will become a powerful vertebrate model system to study the role of subtle genotypic variation in phenotypic diversity.

Will the zebrafish genome sequence lead to new concepts in biology? Quite possibly. It is now obvious that the full impact of the human genome sequence was not apparent upon its release. For example, who could have predicted the discovery of thousands of RNA molecules that do not encode proteins but which are now known to have key regulatory functions, or foreseen the progress in using genome sequences to reconstruct the history of human evolution¹²? The zebrafish genome sequence and mutant collection might just be the first steps on another avenue of discovery. ■

Alexander F. Schier is in the Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA.
e-mail: schier@mcb.harvard.edu

1. Howe, K. *et al.* *Nature* **496**, 498–503 (2013).
2. Kettleborough, R. N. W. *et al.* *Nature* **496**, 494–497 (2013).
3. Varshney, G. K. *et al.* *Genome Res.* **23**, 727–735 (2013).
4. Hwang, W. Y. *et al.* *Nature Biotechnol.* **31**, 227–229 (2013).
5. Chang, N. *et al.* *Cell Res.* **23**, 465–472 (2013).
6. Zu, Y. *et al.* *Nature Meth.* **10**, 329–331 (2013).
7. Bedell, V. M. *et al.* *Nature* **491**, 114–118 (2012).
8. Driever, W. *et al.* *Development* **123**, 37–46 (1996).
9. Haffter, P. *et al.* *Development* **123**, 1–36 (1996).
10. Solnica-Krezel, L., Schier, A. F. & Driever, W. *Genetics* **136**, 1401–1420 (1994).
11. Zon, L. I. & Peterson, R. T. *Nature Rev. Drug Discov.* **4**, 35–44 (2005).
12. Lander, E. S. *Nature* **470**, 187–197 (2011).

Macrophage biology in development, homeostasis and disease

Thomas A. Wynn¹, Ajay Chawla² & Jeffrey W. Pollard^{3,4}

Macrophages, the most plastic cells of the haematopoietic system, are found in all tissues and show great functional diversity. They have roles in development, homeostasis, tissue repair and immunity. Although tissue macrophages are anatomically distinct from one another, and have different transcriptional profiles and functional capabilities, they are all required for the maintenance of homeostasis. However, these reparative and homeostatic functions can be subverted by chronic insults, resulting in a causal association of macrophages with disease states. In this Review, we discuss how macrophages regulate normal physiology and development, and provide several examples of their pathophysiological roles in disease. We define the ‘hallmarks’ of macrophages according to the states that they adopt during the performance of their various roles, taking into account new insights into the diversity of their lineages, identities and regulation. It is essential to understand this diversity because macrophages have emerged as important therapeutic targets in many human diseases.

Macrophages, which were originally identified by Metchnikoff on account of their phagocytic nature, are ancient cells in metazoan phylogeny. In adult mammals, they are found in all tissues where they display great anatomical and functional diversity. In tissues, they are organized in defined patterns with each cell occupying its own territory, a type of tissue within a tissue. Although several attempts have been made to classify macrophages, the most successful definition is the mononuclear phagocytic system (MPS), which encompasses these highly phagocytic cells (professional phagocytes) and their bone marrow progenitors. In the MPS schema, adult tissue macrophages are defined as end cells of the mononuclear phagocytic lineage derived from circulating monocytes that originate in the bone marrow. However, this definition is inadequate as macrophages have several origins during ontogeny and each of these different lineages persist into adulthood¹. Other functional classifications of macrophages have included binary classifications that refer to inflammatory states. These include the activated macrophage and alternatively activated macrophage (AAM) categories, and the derivative M1 and M2 categories for these types of macrophage in the non-pathogen-driven condition^{2,3}. These two states are defined by responses to the cytokine interferon- γ (IFN- γ) and activation of Toll-like receptors (TLRs), and to interleukin-4 (IL-4) and IL-13, respectively. Although this classification is a useful heuristic that may reflect extreme states, such as that of activated macrophages during immune responses mediated by T helper cells that express IFN- γ (T_H1) or of AAMs during parasitic infections², such binary classifications cannot represent the complex *in vivo* environment for most macrophage types, in which numerous other cytokines and growth factors interact to define the final differentiated state. Indeed, transcriptional profiling of resident macrophages by the Immunological Genome Project show that these populations have high transcriptional diversity with minimal overlap, suggesting that there are many unique classes of macrophages⁴.

Macrophages have roles in almost every aspect of an organism's biology; from development, homeostasis and repair, to immune responses to pathogens. Resident macrophages regulate tissue homeostasis by acting as sentinels and responding to changes in physiology as well as

challenges from outside. During these homeostatic adaptations, macrophages of different phenotypes can also be recruited from the monocyte reservoirs of blood, spleen and bone marrow⁴, and perhaps from resident tissue progenitors or through local proliferation^{5,6}. Unfortunately, in many cases these homeostatic and reparative functions can be subverted by continuous insult, resulting in a causal association of macrophages with disease states, such as fibrosis, obesity and cancer (Fig. 1). Thus, macrophages are an incredibly diverse set of cells that constantly shift their functional state to new metastable states (‘set points’) in response to changes in tissue physiology or environmental challenges. They should not even be considered as one cell type but should be subdivided into different functional subsets according to their different origins.

Macrophage responses to pathogens have been discussed previously^{2,7,8} and therefore this Review focuses on the homeostatic mechanisms by which macrophages contribute to physiological and pathophysiological adaptations in mammals. Here we define the hallmarks of macrophages that perform particular functions, taking into account new insights into the diversity of their lineages, identity and regulation. This phenotypic diversity is essential to understand because macrophages are central to many disease states and have emerged as important therapeutic targets in many diseases.

Macrophage origins rewritten

Ontologically, the MPS has been proposed to arise from a rigid temporal succession of macrophage progenitors⁹. In mice, these start to develop first at embryonic day 8 from the primitive ectoderm of the yolk sac and give rise to macrophages that do not have a monocytic progenitor. This primitive system is followed by definitive haematopoiesis in the fetal liver, which is initially seeded by haematopoietic progenitors from the yolk sac and subsequently from the hematogenic endothelium of the aorto-gonadal-mesonephros region of the embryo. After this point, the fetal liver is the source of definitive haematopoiesis that generates circulating monocytes during embryogenesis. Coincident with the postnatal formation of bone, fetal liver haematopoiesis declines and is replaced by bone marrow haematopoiesis. This definitive haematopoiesis is the source of

¹Immunopathogenesis Section, Program in Tissue Immunity and Repair and Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20877-8003, USA. ²Cardiovascular Research Institute, Department of Physiology and Medicine, University of California San Francisco, California 94158-9001, USA. ³Medical Research Council Centre for Reproductive Health, Queen's Medical Research Institute, University of Edinburgh, Edinburgh EH16 4TJ, UK. ⁴Center for the Study of Reproductive Biology and Women's Health, Department of Developmental and Molecular Biology, Albert Einstein College of Medicine, New York, New York 10461, USA.

Normal physiology	Pathology
Microglia, (neuronal patterning, fluid balance)	Neurodegeneration
Osteoclasts and macrophages (bone remodelling; haematopoiesis)	Osteoporosis and osteopetrosis Leukemia
Heart and vasculature	Atherosclerosis
Kupffer cells (lipid metabolism, toxin removal)	Fibrosis
Branching morphogenesis	Cancer and metastasis
Metabolism; adipogenesis	Obesity and diabetes
Immunity	Arthritis, EAE, IBD

Figure 1 | Macrophages in development, homeostasis and disease.

Macrophages have many developmental roles in shaping the architecture of various tissues, such as brain, bone and mammary gland tissues. After development of the organism, macrophages modulate homeostasis and normal physiology through their regulation of diverse activities, including metabolism and neural connectivity, and by detecting damage. However, these trophic and regulatory roles are often subverted by continuous insult, and macrophages contribute to many diseases that are often associated with ageing. EAE, experimental autoimmune encephalomyelitis; IBD, inflammatory bowel disease.

circulating monocytes (resident, lymphocyte antigen 6c negative (Ly6c⁻) and inflammatory Ly6c⁺ in mice) and from which it has been considered that all resident macrophages in tissues are derived⁴. However, this model for the formation of the MPS has been challenged (Fig. 2). First, lineage-tracing experiments have shown that microglia are primarily derived from the yolk-sac progenitors, whereas Langerhans cells have a mixed origin from yolk sac and fetal liver^{10,11}. Second, experiments using ablation of c-Myb-dependent bone marrow haematopoiesis followed by transplantation with genetically dissimilar bone marrow together with lineage tracing showed that the major tissue-resident population of macrophages (defined as F4/80 bright) in skin, spleen, pancreas, liver, brain and lung arise from yolk sac progenitors. In a few tissues, such as kidney and lung, macrophages have a chimaeric origin being derived from yolk sac (F4/80^{high}) and bone marrow (F4/80^{low}). In contrast to this yolk sac and fetal liver origin for most macrophages, classical dendritic cells and the F4/80^{low} macrophages are continuously replaced by bone-marrow-derived progenitors⁶. These data indicate that there are at least three lineages of macrophages in the mouse, which arise at different stages of development and persist to adulthood. The data also call into question the function of circulating monocytes because, at least in mice, these cells do not seed the majority of the adult tissues with macrophages. In fact, complete loss of CD16⁺ monocytes in humans seems to be of little consequence¹². Thus, the function of monocytes needs to be defined with the possibility that patrolling monocytes (Ly6c⁻) act to maintain vessel integrity and to detect pathogens while inflammatory monocytes (Ly6c⁺) are recruited predominantly to sites of infection or injury, or to tissues that have continuous cyclical recruitment of macrophages, such as the uterus.

Regardless of their origin, genetic and cell culture studies indicate that the major lineage regulator of almost all macrophages is macrophage colony-stimulating factor 1 receptor (CSF1R). This class III transmembrane

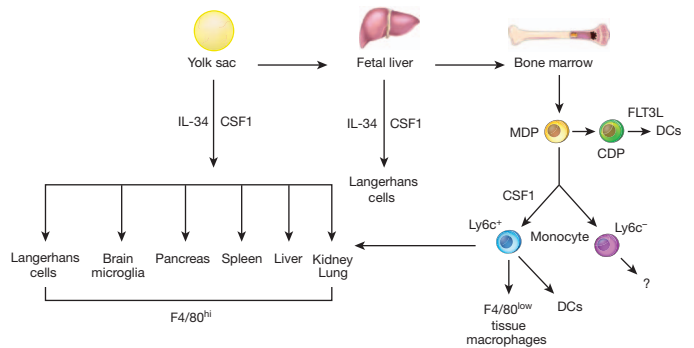


Figure 2 | A redefined model of macrophage lineages in mice. The mononuclear phagocytic system in adults derives from at least three sources. The first is the yolk sac, which produces progenitors that populate all tissues and that have progeny that persist throughout life as F4/80 bright resident macrophages. These lineages are mainly regulated by CSF1R and its ligands, IL-34 and CSF1. The second is the fetal liver, and this is less well defined but seems to contribute to the production of adult Langerhans cells, perhaps through a progenitor that is derived from the yolk sac. The third lineage derives from the bone marrow (BM) to give circulating monocytes and their progeny F4/80^{low} macrophages, and dendritic cells (DCs). In this case the Ly6c⁺ monocytes give rise to the classic Steinman dendritic cells under the regulation of FLT3, and these are continuously replenished. Other macrophages that are F4/80^{low} also emanate from Ly6c⁺ monocytes, and in some cases—such as in kidney and lung—they co-exist with those derived from the yolk sac to give chimaeric organs. The exact role of the patrolling Ly6c⁻ macrophages, and the contribution of fetal liver to adult tissue macrophages, remain unclear. CDP, committed dendritic cell progenitor; MDP, monocyte dendritic cell progenitor.

tyrosine kinase receptor is expressed on most, if not all, mononuclear phagocytic cells, and a reporter mouse expressing green fluorescent protein (GFP) from the *Csf1r* locus illustrates their relative abundance (5–20% of cells) and tissue distribution¹³. *Csf1r* expression and its requirement for differentiation distinguish macrophages from many, but not all, dendritic-cell subtypes¹⁴. Targeted ablation of the *Csf1r* causes severe depletion of macrophages in many tissues, such as brain, skin, bone, testis and ovary. Moreover, an initial comparison of the *Csf1r*-null mice with those homozygous for a spontaneous (osteopetrotic (*Csf1*^{op})) null mutation in its cognate ligand (*Csf1*^{op/op} mice) demonstrated that all phenotypes in the *Csf1r*-null mice were also found in the *Csf1*^{op/op} mice, indicating that CSF1 has only a single receptor¹⁵. However, the phenotype of the *Csf1r*-null mice is more severe than that of the *Csf1*-null mice, including the complete loss of microglia and Langerhans cells^{10,16} in the *Csf1r*-null mice, which suggested the presence of another ligand. Indeed, IL-34, with a distinct but overlapping pattern of expression with *Csf1*, was recently identified as an additional ligand for the CSF1R¹⁷. Targeted ablation of *Il34* resulted in loss of microglia and Langerhans cells, but had little impact on bone marrow, liver or splenic macrophages¹⁸.

Despite the importance of the CSF1R in macrophage specification, *Csf1r*-null mutant mice still have some tissue macrophages, such as in the spleen, indicating the existence of other macrophage growth factors. Potential candidates include granulocyte-macrophage colony-stimulating factor (GM-CSF) and IL-3, which act as macrophage growth factors in tissue culture. However, mice lacking GM-CSF or IL-3 do not show notable defects in their tissue macrophages, except in alveolar macrophages, which indicates that they are regulated by GM-CSF¹⁹. Vascular endothelium growth factor A (VEGFA) proteins are another candidate regulator of macrophages because they can compensate for the loss of *Csf1* in osteoclast development *in vivo*²⁰. In contrast to CSF1 that is found in all tissues and serum, and is a basal regulator of macrophage number through a negative feedback loop¹⁵, GM-CSF is not a steady-state ligand and seems to be synthesized in response to challenge²¹. GM-CSF and FLT3L regulate the maturation of dendritic cell populations with the notable exception of Langerhans cells, whose development is dependent on *Csf1r*²². Recent genomic profiling of Langerhans cells place them closer to macrophages than dendritic cells, and this data together with

their lineage dependence on *Csf1r* may indicate that classification should be updated¹⁴. Dendritic cells will not be discussed further in this Review, but their biology and lineages have been extensively reviewed recently¹⁴.

In their basal state, resident tissue macrophages show great diversity in their morphologies, transcriptional profiles, anatomical locations and functional capabilities²³. This functional heterogeneity probably results from the dynamic crosstalk between resident tissue macrophages and the client cells that they support. To understand this macrophage diversity there must be an understanding of transcriptional regulation. The most important of these transcription factors is SFPI1 (also known as PU.1), a member of the ETS family whose loss following targeted mutation results in complete depletion of CD11b⁺F4/80⁺ macrophages, including those derived from the yolk sac⁶. However, *Sfp1* action is not limited to macrophages as B cells are also severely depleted in these *Sfp1*-null mutant mice. Similarly, other members of the ETS family are also involved in macrophage differentiation, including *Ets2*, which positively regulates the *Csf1r* promoter. In adults, *Mafk* (also known as *v-Maf*) is required for the local proliferation that maintains resident macrophages⁴. In the differentiation of osteoclasts, *Fos* and *Mitf* are required²⁴, whereas *Gata2* is required for monocyte development but not for resident macrophage populations²⁵. However, little is known about the transcriptional control of the differentiation of the diverse tissue macrophages, such as those in the liver and brain¹³. Most research has focused on their functional activation in response to environmental challenges²³, as discussed below. Nevertheless, the recent transcriptional profiling of resident macrophages has identified many candidate transcription factors, including those that may regulate core macrophage-associated genes such as *Mitf* (microphthalmia) family members, *Tcf3*, *Cebpa*, *Bach1*, *Creg1* and genes that are unique to subpopulations, including *Gata6* and *Spic*, whose targeted gene ablation will undoubtedly define subsets of macrophages and their unique activities¹.

Macrophages in development

Metchnikoff proposed that macrophages participate in the maintenance of tissue integrity and homeostasis. To do so, macrophages would need to be able to discriminate self from non-self, sense tissue damage and recognize invading pathogens, an insight that led to the concept of innate immunity for which he was awarded the Nobel prize. The inherent properties of macrophages, which include sensing inside from out, motility throughout the organism, phagocytosis and degradation, were later sequestered to instruct the acquired immune system as it evolved to more efficiently deal with changing pathogenic challenges. This enhanced sophistication of the immune system probably resulted in the evolution of dendritic cells as specialized mononuclear phagocytes to interface with the acquired immune system. Indeed, in mammals, dendritic cells seem to be focused on initiating tissue immune responses, whereas tissue macrophages seem to be focused on homeostasis and tissue integrity⁹.

Emphasis on the immunological and repair aspects of macrophage function has overshadowed their importance in the development of many tissues; for example, studies of *Csf1*^{op/op} mice, which lack many macrophage populations, have revealed a cluster of developmental abnormalities¹⁹. Most notable among these is the development of osteopetrosis, which is caused by the loss of bone-reabsorbing macrophages known as osteoclasts. This phenotype, which is also observed in *Sfp1*-null mice, is axiomatic for the roles of macrophages in development, in that cell fate decisions are unchanged but the tissue remodelling and expression of growth factors is lost. Specifically, although bone formation is intact in *Csf1*- or *Sfp1*-null mice, the bones are not sculpted to form the cavities in which haematopoiesis commences¹⁹. Consequently, the functional integrity of the bones, in terms of load bearing and haematopoiesis, is compromised. *Csf1*^{op/op} mice survive to adulthood because of extra-medullary haematopoiesis in the spleen and liver¹⁹, and as mice age, osteoclastogenesis is rescued by compensatory expression of VEGF and therefore bone marrow haematopoiesis commences²⁰.

Remodelling deficiencies in the absence of macrophages have also been noted in several other tissues, including the mammary gland, kidney and pancreas, suggesting a general requirement for macrophages in tissue patterning and branching morphogenesis^{19,26}. In the mammary gland, the best studied of these tissues, macrophages are recruited to the growing ductal structure and their loss results in a slower rate of outgrowth and limited branching, phenotypes that are reiterated during the mammary growth caused by pregnancy¹⁹. This stems partly from the failure to remodel the extracellular matrix during the outgrowth of the ductal structures. However, recent studies have also implicated macrophages in maintaining the viability and function of mammary stem cells, which reside at the tip of the duct known as the terminal end bud and are responsible for the outgrowth of this structure²⁷. In stem cell biology similar roles for macrophages have been suggested in the maintenance of intestinal integrity and its regeneration after damage²⁸, whereas a subpopulation of macrophages in the haematopoietic niche regulates the dynamics of haematopoietic stem cell release and differentiation²⁹. Furthermore, in regenerating livers, macrophages specify hepatic progenitor fate through the expression of WNT ligands and antagonism of Notch signalling³⁰. Macrophage control of stem cell function is clearly an emerging and important research area.

As 'professional' phagocytes (macrophages were originally defined by their exceptional phagocytic ability), macrophages perform critical functions in the remodelling of tissues, both during development and in the adult animal; for example, during erythropoiesis, maturing erythroblasts are surrounded by macrophages that ingest the extruded erythrocyte nuclei. Remarkably, this function of macrophages is critical because in its absence, erythropoiesis is blocked and lethality ensues³¹. Macrophages also make decisions about haematopoietic egress from the bone marrow through engulfing cells that do not express the CD47 ligand³². They also maintain the haematopoietic steady state through engulfment of neutrophils and erythrocytes in the spleen and liver, and the failure of this activity results in neutropenia, splenomegaly and reduced body weight³³. Phagocytosis, particularly of apoptotic cells, is clearly central to macrophage function and this is emphasized by the build-up in macrophage-depleted mice of such cells during development; for example, during the resolution of the inter-digit areas during limb formation³⁴. However, there is no apparent consequence to this phenomenon, as less-efficient 'non-professional' phagocytes clear excess apoptotic cells. Despite this, macrophages have evolved to 'eat' cells, and to suppress inflammation and autoimmunity in response to self-antigens that may arise during homeostasis³⁵.

Macrophages also regulate angiogenesis through a number of mechanisms. This has been most extensively studied in the eye during its development. Early in the postnatal period, during regression of the hyaloid vasculature, macrophages identify and instruct vascular endothelial cells to undergo apoptosis if these cells do not receive a counterbalancing signal from pericytes to survive. WNT7B that is synthesized by macrophages delivers this cell-death signal to the vascular endothelial cells, and in the absence of either WNT7B or macrophages there is vascular over-growth³⁶. WNT secretion is also required later in retinal vasculature development but in this case macrophage synthesized WNT5A and WNT11, a non-canonical WNT, induces expression of soluble VEGF receptor 1 (VEGFR1) through an autocrine mechanism that titrates VEGF and thereby reduces vascular complexity so that the vascular system is appropriately patterned³⁷. Furthermore, at other times of ocular development, macrophages regulate vascular complexity. In this circumstance, macrophage-synthesized VEGFC reinforces Notch signalling³⁸. In addition, during angiogenesis in the hindbrain, macrophages enhance the anastomosis of tip and stalk cells to give functional vessels³⁹. These macrophage functions are not restricted to the vascular arm of the circulatory system, as they also have roles in lymphangiogenesis during development⁴⁰, and in adults they have a notable role in maintaining fluid balance through their synthesis of VEGFC⁴¹.

Brain development is also influenced by macrophages. These macrophages called microglia depend on CSF1R signalling for their presence^{10,16}.

In the absence of this signalling there are no microglia, and the brains of these mice have substantial structural defects as they mature¹⁶. Both CSF1 and IL-34 are expressed by neurons in a mutually restricted pattern of expression, and IL-34 is the major factor for microglial differentiation and viability^{10,42}. The disruption of architecture in the brain of the *Csf1r*-null mouse, together with well-documented deficiencies in neuronal processing regulating olfaction and the reproductive axis in the hypothalamus in *Csf1*-null mice, strongly suggests that microglia are involved in the development of neuronal circuitry and the maintenance of brain structure^{16,19}. Indeed, microglia have been shown to promote neuron viability¹⁹, modulate neuronal activity⁴³ and prune synapses during development⁴⁴, as well as express a range of neuronal growth and survival factors, including NGF¹⁹. This conjecture is supported by the finding that hypomorphic mutation in *CSF1R* in humans is responsible for hereditary diffuse leukoencephalopathy with spheroids that results from loss of myelin sheaves and axonal destruction⁴⁵. These trophic activities of microglia are also consistent with macrophages having roles in neuroprotection after injury, as defined in a variety of models. These effects include the promotion of survival and proliferation of retinal progenitor cells, and the regeneration of adult sensory neurons^{46–48}. However, caution needs to be exercised in attributing all of the phenotypes observed in the brains of *Csf1r*-mutant mice or humans to the loss of microglia, as *Csf1r* expression has been reported on neuronal stem cells and their development *in vivo* is regulated by CSF1R⁴². Nevertheless, it seems likely that microglia have important roles in the development of neuronal circuitry, though their effects on the proliferation, survival and connectivity of neurons⁴³, through their effects on myelination, or by modulating angiogenesis and fluid balance in the brain¹⁶.

The examples given above indicate a few of the roles for macrophages in normal development and these are likely to expand with further study. Phenotypically in mice, macrophages are CD11b⁺, CD68⁺ CSF1R⁺ F4/80⁺ and phagocytic and their activities are through the temporal and spatial delivery of developmentally important molecules such as VEGFs and WNTs as well as proteases. These developmental roles of macrophages are re-capitulated in repair as described below but are also intimately involved in chronic conditions that lead to pathologies as well as the development and progression of malignancies.

Macrophages in metabolic homeostasis

Mammalian metabolic organs, such as the liver, pancreas and adipose tissue, are composed of parenchymal and stromal cells, including macrophages, which function together to maintain metabolic homeostasis⁴⁹. By regulating this interaction, mammals are able to make marked adaptations to changes in their environment and in nutrient availability. For example, during bacterial infection, innate activation of macrophages results in secretion of pro-inflammatory cytokines, such as TNF- α , IL-6 and IL-1 β , which collectively promote peripheral insulin resistance to decrease nutrient storage^{50,51}. This metabolic adaptation is necessary for mounting an effective defence against bacterial and viral pathogens because nearly all activated immune cells preferentially use glycolysis to fuel their functions in host defence. However, this adaptive strategy of nutrient re-allocation becomes maladaptive in the setting of diet-induced obesity, a state that is characterized by chronic low-grade macrophage-mediated inflammation^{51,52}. In the sections below, we provide a general framework for understanding the pleiotropic functions carried out by macrophages to maintain metabolic homeostasis (Fig. 3). Although our current knowledge in this area is primarily derived from studies in obese insulin-resistant mice, it is likely that tissue-resident macrophages also participate in facilitating metabolic adaptations in healthy animals.

White adipose tissue

White adipose tissue (WAT) is not only the principal site for long-term storage of nutrients but also regulates systemic metabolism through the release of hormones called adipokines⁵³. These metabolic functions of WAT are primarily performed by adipocytes with trophic support provided by stromal cells, including macrophages. Thus, macrophage

representation in WAT, both in terms of numbers and their activation state, reflects the metabolic health of adipocytes⁵¹. For example, in lean healthy animals, adipose tissue macrophages comprise 10–15% of stromal cells and express the canonical markers (Arg1⁺, CD206⁺, CD301⁺) of AAMs⁵⁴. In contrast, macrophage content increases to 45–60% during obesity^{55,56}, resulting from increased recruitment of Ly6C^{hi} monocytes that differentiate into inflammatory macrophages, as judged by their expression of *Nos2*, *Tnfa* (also known as *Tnf*) and *Itgax*^{54,55}. Although these macrophages contribute to the development of insulin resistance in adipocytes, recent studies suggest that these cells also participate in remodelling of the enlarging WAT, functions that facilitate the storage of excess nutrients in adipocytes⁵⁷. This suggests that two macrophage subsets coordinate homeostatic adaptations in adipocytes of lean and obese animals.

In healthy animals, AAMs are critical for maintaining insulin sensitivity in adipocytes⁵¹. This trophic effect of AAMs is partly mediated by secretion of IL-10, which potentiates insulin action in adipocytes⁵⁴. These observations led various groups to focus on cell-intrinsic and cell-extrinsic mechanisms that control alternative activation of adipose tissue macrophages. For cell-intrinsic factors, transcription factors downstream of IL-4 and IL-13 signalling, such as PPAR- γ , PPAR- δ and KLF4, were found to be required for the maintenance of AAMs in WAT and metabolic homeostasis^{58–61}. The dominant cell-extrinsic factors regulating maturation of AAMs in lean WAT are the type 2 cytokines IL-4 and IL-13 (ref. 60). Absence of eosinophils, which constitute the major cell type capable of IL-4 secretion in WAT⁶², impairs alternative activation of adipose tissue macrophages and makes mice susceptible to obesity-induced insulin resistance. Together, these reports have established that homeostatic functions performed by AAMs in WAT are required for metabolic adaptations to excessive nutrient intake.

Although adipocytes in lean animals can easily accommodate acute changes in energy intake, chronic increase in energy intake places adipocytes under considerable metabolic stress. Consequently, the enlarging WAT releases chemokines, such as CC-chemokine ligand 2 (CCL2), CCL5 and CCL8, to recruit Ly6C^{hi} inflammatory monocytes into the WAT⁶³, where these cells differentiate into CD11c⁺ macrophages and form 'crown-like structures' around dead adipocytes^{54,64}. As these CD11c⁺ macrophages phagocytize dead adipocytes and become lipid engorged, they initiate expression of inflammatory cytokines, such as TNF- α and IL-6, which promote insulin resistance in the surrounding adipocytes⁵⁴. Presumably, this initial decrease in adipocyte insulin sensitivity is an adaptation to limit nutrient storage. However, in the setting of unabated increase in caloric intake, this adaptive response becomes maladaptive, contributing to pathogenesis of obesity-induced systemic insulin resistance.

Brown adipose tissue

In mammals, brown adipose tissue (BAT) is the primary thermogenic organ that is activated by exposure to environmental cold⁶⁵. For decades, it had been thought that hypothalamic sensing of cold triggers an increase in sympathetic nerve activity to stimulate the BAT program of adaptive thermogenesis⁶⁵. However, recent work has demonstrated that resident macrophages are required to facilitate the metabolic adaptations of BAT and WAT to cold. Specifically, exposure to cold temperatures results in alternative activation of BAT and WAT macrophages, which are required for induction of thermogenic genes in BAT and lipolysis of stored triglycerides in WAT⁶⁶. Accordingly, mice lacking AAMs are unable to mobilize fatty acids from WAT to maximally support BAT thermogenesis, which is necessary for the maintenance of core body temperature in cold environments. These supportive functions of macrophages are mediated by their secretion of norepinephrine, which surprisingly accounts for approximately 50% of the catecholamine content of BAT and WAT in the cold. Thus, cold-induced alternative activation of BAT and WAT macrophages provides an example of how resident macrophages provide trophic support to facilitate the

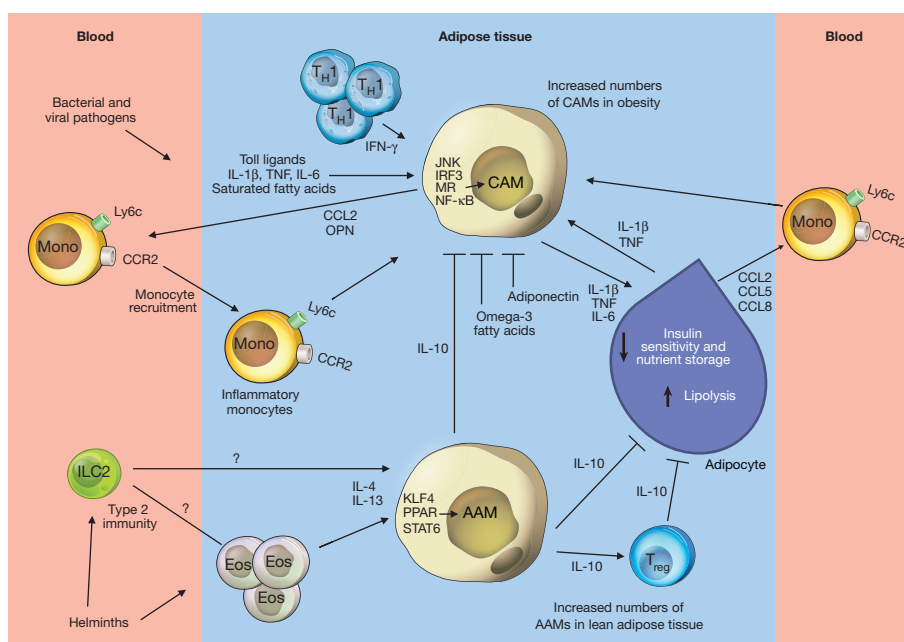


Figure 3 | Activated and alternatively activated macrophages differentially regulate insulin sensitivity in obesity. In lean healthy animals, adipose tissue macrophages comprise 10–15% of stromal cells, and express markers that link them with AAMs, which are critical for maintaining insulin sensitivity in adipocytes, partly through the production of IL-10. Type 2 cytokines such as IL-4 and IL-13, which are derived from a variety of cellular sources, including eosinophils, seem to be important for the maintenance of the AAM phenotype in lean tissues. In contrast, during obesity, Ly6c^{hi} monocytes are recruited, which increases macrophage content to 45–60%. These macrophages, in contrast to normal resident macrophages, express an inflammatory phenotype,

characterized by the production of TNF- α , IL-6 and IL-1 β . These inflammatory macrophages decrease insulin sensitivity while facilitating the storage of excess nutrients. The enlarging white adipose tissues in turn release chemokines, such as CCL2, CCCL5 and CCL8, to recruit additional Ly6c^{hi} inflammatory monocytes that exacerbate the process. This mechanism is also enhanced during bacterial and viral infections, so essential nutrients are diverted to lymphocytes, which must use glycolysis to enhance their activation at times of stress. CAM, classically activated macrophage. Eos, eosinophils; ILC2, type 2 innate lymphoid cells; Mono, monocytes.

function of tissue parenchymal cells, in this case the white and brown adipocytes.

Liver and pancreas

Liver integrates nutrient, hormonal and environmental signals to maintain glucose and lipid homeostasis in mammals. Over the past few years, evidence has emerged that Kupffer cells, the resident macrophages of liver, facilitate the metabolic adaptations of hepatocytes during increased caloric intake. During obesity, an imbalance between the uptake, synthesis and oxidation of fatty acids results in increased lipid storage in hepatocytes, a key factor in the development of hepatic insulin resistance⁶⁷. Interestingly, Kupffer cells directly participate in this process by regulating the oxidation of fatty acids in hepatocytes. An early insight into this process came from studies that identified PPAR- δ as an important regulator of the IL-4- and IL-13-driven program of alternative macrophage activation^{58,61}. These studies revealed that loss of PPAR- δ in myeloid cells specifically impaired alternative activation of Kupffer cells, resulting in hepatic steatosis and insulin resistance. A similar phenotype was observed when Kupffer cells were depleted in rodents using gadolinium chloride or clodronate-containing liposomes⁶⁸. Although the precise factors elaborated by Kupffer cells are still not known, co-culture studies suggest that Kupffer-cell-derived factors work in a *trans*-acting manner to maintain hepatic lipid homeostasis^{58,61}.

Pancreas functions as an endocrine and exocrine gland in mammals. Recent findings suggest that, analogous to obesity-induced WAT inflammation, high-fat feeding induces the infiltration of macrophages into the insulin-producing islets. In this case, the increased intake of dietary lipids results in beta-cell dysfunction, which induces the expression of chemokines, such as CCL2 and CXCL1, to recruit inflammatory monocytes or macrophages into the islets^{69,70}. Consequently, the secretion of IL-1 β and TNF- α by the infiltrating macrophages augments

beta-cell dysfunction, resulting in impaired insulin secretion and hyperglycaemia in obese mice. Although these reports have elucidated the pathogenic role of macrophages in beta-cell dysfunction, in the future it will be important to determine whether macrophages also participate in the physiological regulation of beta-cell biology as they do during development and pregnancy¹⁹.

Macrophages in disease

When tissues are damaged following infection or injury, inflammatory monocytes (Ly6c⁺ in mice) are recruited from the circulation and differentiate into macrophages as they migrate into the affected tissues⁴. These recruited macrophages often show a pro-inflammatory phenotype in the early stages of a wound-healing response. They secrete a variety of inflammatory mediators, including TNF- α , IL-1 and nitric oxide, which activate anti-microbial defence mechanisms, including oxidative processes that contribute to the killing of invading organisms⁷. They also produce IL-12 and IL-23, which direct the differentiation and expansion of anti-microbial T_H1 and T_H17 cells (T helper cells that express IFN- γ and IL-17) that help to drive inflammatory responses forward³. Although these inflammatory macrophages are initially beneficial because they facilitate the clearance of invading organisms, they also trigger substantial collateral tissue damage because of the toxic activity of reactive oxygen and nitrogen species and of T_H1 and T_H17 cells⁷¹. Indeed, if the inflammatory macrophage response is not quickly controlled, it can become pathogenic and contribute to disease progression, as is seen in many chronic inflammatory and autoimmune diseases^{72,73}. To counteract the tissue-damaging potential of the inflammatory macrophage response, macrophages undergo apoptosis or switch into an anti-inflammatory or suppressive phenotype that dampens the pro-inflammatory response while facilitating wound healing⁷. These regulatory macrophages often produce ligands associated with development,

such as WNT ligands, that are essential for tissue repair⁷⁴. It is becoming increasingly clear that the mechanisms that regulate the transformation of inflammatory macrophages into an anti-inflammatory cell or suppressive macrophages back into a pro-inflammatory phenotype has a major impact on the progression and resolution of many chronic diseases, as discussed below (Fig. 4).

Macrophages in cancer

Tumours are abundantly populated by macrophages³. Although macrophages were originally thought to be part of an anti-tumour response, clinical and experimental data indicate that in the large majority of cases macrophages promote tumour initiation, progression and metastasis⁷⁵. In response to persistent infections or chronic irritation, macrophages synthesize inflammatory cytokines, IFN- γ , TNF- α and IL-6, which engage other immune cells to sustain the chronic inflammation that seems to be causal in tumour initiation and promotion⁷⁶. The tumour-inducing activities are multi-factorial; for example, through the production of inflammatory cytokines, such as IFN- γ in skin cancer that is induced by exposure to ultraviolet light⁷⁷ and TNF- α in carcinogen-induced cancer, through the generation of a mutagenic environment^{76,78} or through alterations of the microbiome⁷⁹. However, once tumours become established they cause differentiation so that the tumour-associated macrophages (TAMs) change from an immunologically active state to adopt a trophic immunosuppressive phenotype that promotes tumour progression and malignancy (they become 'tumour-educated')⁷⁵.

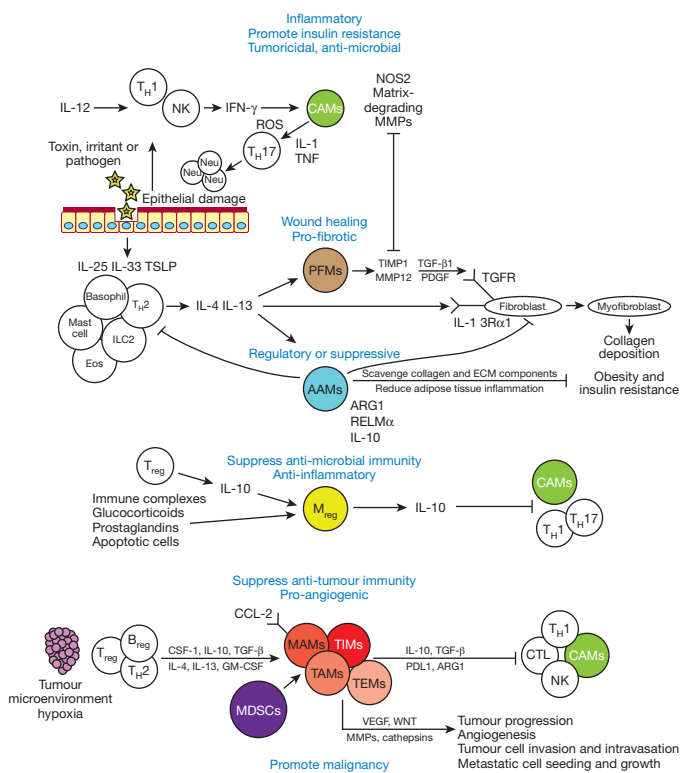
In established tumours, TAMs stimulate tumour-cell migration, invasion and intravasation, as well as the angiogenic response required for tumour growth^{75,80,81}. These events are required for tumour cells to become metastatic, as they facilitate their escape into the circulatory or lymphatic system. Evidence from autochthonous models of breast cancer suggests that the macrophages take on these activities in response to CSF1, IL-4 and IL-13 encountered in the tumour microenvironment. For example, IL-4-mediated differentiation⁸⁰ results in a reciprocal paracrine dialogue between CSF1 and EGF, synthesized by tumour cells and TAMs, respectively, that promotes tumour-cell invasion and intravasation in mammary cancer⁸². In mammary cancers, this loop is initiated by CXCL12 in the polyoma virus middle T (PyMT) model or

heregulin (also known as pro-neuregulin-1, membrane-bound isoform) in the HER2/Neu model. In human xenograft models, CCL18 is also required for tumour-cell invasion and metastasis, because it has a role in triggering integrin clustering⁸³. TAMs also remodel the tumour microenvironment through the expression of proteases such as matrix metalloproteinases (MMPs), cathepsins and urokinase plasminogen activator, and matrix remodelling enzymes such as lysyl oxidase and SPARC^{81,84}. The proteases, such as cathepsin B, MMP2, MMP7 and MMP9, cleave extracellular matrix and thereby provide conduits for the tumour cells and release growth factors such as heparin-binding EGF (HB-EGF) and EGF mimics that foster tumour-cell invasion and metastasis^{84,85}.

Macrophages have an important role in tumour angiogenesis as they regulate the marked increase in vascular density, known as the angiogenic switch, that is required for the transition to the malignant state⁸⁶. These angiogenic TAMs are characterized by the expression of the angiopoietin receptor TIE2, which is also expressed in macrophages during development^{87,88}. Ablation of this specific population inhibits tumour angiogenesis and thus tumour growth and metastasis in a variety of models^{87,88}. TAMs secrete many angiogenic molecules, including VEGF family members TNF- α , IL-1 β , IL-8, PDGF and FGF^{75,88,89}. Of these, myeloid-derived VEGF is required for the angiogenic switch⁸⁹ but other aspects of angiogenesis can be independent of VEGF and involve the secreted protein Bv8 (also known as prokineticin 2 or

Figure 4 | Macrophages that exhibit unique activation profiles regulate disease progression and resolution.

Macrophages are highly plastic cells that adopt a variety of activation states (different coloured circles) in response to stimuli that are found in the local environment. During pathogen invasion or after tissue injury or exposure to environmental irritants, local tissue macrophages often adopt an activated or 'inflammatory phenotype'. These cells are commonly called classically activated macrophages (CAMs), because they were the first activated macrophage population to be formally defined. These macrophages are activated by IFN- γ and/or after TLR engagement, leading to the activation of the NF- κ B and STAT1 signalling pathways. This in turn increases the production of reactive oxygen and nitrogen species, and pro-inflammatory cytokines, like TNF- α , IL-1 and IL-6, that enhance anti-microbial and anti-tumour immunity, but may also contribute to the development of insulin resistance and diet-induced obesity. In contrast, some epithelium-derived alarmins and the type 2 cytokines IL-4 and IL-13 result in an 'alternative' state of macrophage activation (AAMs) that has been associated with wound healing, fibrosis, insulin sensitivity and immunoregulatory functions. They also activate wound-healing, pro-angiogenic and pro-fibrotic macrophages (PFMs) that express TGF- β 1, PDGF, VEGF, WNT ligands, and various matrix metalloproteinases that regulate myofibroblast activation and the deposition of extracellular matrix components. AAMs also express a variety of immunoregulatory proteins, like arginase 1 (ARG1), RELM α , PDL2 and IL-10 that regulate the magnitude and duration of immune responses. These cells also scavenge collagen and extracellular matrix components, and thus the ECM is remodelled. Therefore, in contrast to CAMs that activate immune defenses, AAMs are typically involved in the suppression of immunity and re-establishment of homeostasis. They suppress obesity and insulin resistance that result from the sustained activity of the CAM macrophages. Although type 2 cytokines are important inducers of suppressive or immunoregulatory macrophages, it is now clear that several additional mechanisms can also contribute to the activation of macrophages with immunoregulatory activity. Indeed, IL-10-producing regulatory T (T_{reg}) cells, Fc γ receptor engagement, engulfment of apoptotic cells, and prostaglandins have also been shown to preferentially increase the numbers of regulatory macrophages (M_{reg}) that suppress inflammation and inhibit anti-microbial and anti-tumour defences. The tumour microenvironment itself also promotes the recruitment and activation of immune inhibitory cells, including those of the mononuclear phagocytic series, such as myeloid-derived suppressor cells (MDSCs), tumour-infiltrating macrophages (TIMs), TIE2-expressing macrophages (TEMs), tumour-associated macrophages (TAMs) and metastasis-associated macrophages (MAMs) that promote angiogenesis and tumour growth while suppressing anti-tumour immunity. CTL, cytotoxic T lymphocyte; Neu, neutrophils; NK, natural killer cells; ROS, reactive oxygen species; TSLP, thymic stromal lymphopoietin.



PROK2)⁹⁰. Angiogenic macrophages can be recruited to the tumours by hypoxia^{88,91} but also by growth factors such as CSF1 and VEGF⁹².

Tumours have a proclivity to metastasize to particular sites, and this phenotype is partially defined by macrophages. Data suggest that the tumour-produced fragments of ECM molecules or exosomes prepare these sites, known as pre-metastatic niches, to be receptive to the circulating tumour cells through recruitment of myeloid cells characterized by CD11b and VEGFR1 positivity^{93,94}. These niches are tumour-type-dependent and the fate of the tumour cells can be reprogrammed to a different tissue by the transfer of tumour-conditioned serum to a naive mouse strain⁹³. These niches are also dependent on coagulation as this is necessary for recruitment of the myeloid cells that have recently been more precisely defined as F4/80⁺ monocytes (or F4/80⁺ macrophages)⁹⁵. At lung metastatic sites, mini-clots form that enable the arrest of tumour cells⁹⁵ that then produce CCL2 to recruit CCR2⁺Ly6c⁺ inflammatory monocytes that rapidly develop into Ly6c⁺ metastasis-associated macrophages (MAMs)⁹⁶. These monocytes and MAMs promote tumour-cell extravasation, partly through their expression of VEGF, which induces local vascular permeability. MAMs that are intimately associated with the tumour cells also promote their viability through clustering of tumour-cell-expressed VECAM1 that interlocks with the MAM expressed counter receptor integrin $\alpha 4$ (ref. 83). MAMs also promote subsequent growth of the metastatic cells and, importantly, ablation of these cells after the metastases are established inhibits metastatic growth⁷⁵.

In mice, these individual pro-tumoral functions are carried out by different subpopulations, although they all express canonical markers such as CD11b, F4/80 and CSF1R⁷⁵. This view is consistent with recent profiling of immune cells in various tumour types in mice and humans that indicates that there are differences in the extent of macrophage infiltration and in phenotype⁹⁷. For example, detailed phenotypic profiling in human hepatocellular carcinoma shows various macrophage subtypes defined by specific location that have both pro- and anti-tumoral properties through their engagement of the acquired immune system, although overall the balance is tilted towards pro-tumoral functions⁹⁸. Transcriptional profiling of TAM subpopulations in mice suggest they more closely resemble embryonic macrophages than inflammatory ones, as they have higher expression of developmentally relevant molecules, such as those of the WNT pathway⁷⁵. This strongly suggests that the trophic roles of macrophages found during development, in metabolism and in the maintenance of homeostasis, are subverted by tumours to enhance their growth, invasion and complexity. However, transcriptional control of these different phenotypes is only just being revealed, particularly in *in vivo* contexts³. Many studies have analysed macrophage responses to LPS signalling through nuclear factor- κ B (NF- κ B), but this results in 'activated' macrophages that are mainly involved in antibacterial responses and are likely to be anti-tumoral²³. In contrast, in their trophic and immunosuppressive functions, TAMs are shaped by IL-10 and IL-4 or IL-13 that signal to STAT3 and STAT6, respectively^{3,99}. The PARP proteins and KLF4 also co-operate to induce a pattern of gene expression associated with their tumour-promoting phenotype³. In macrophages, CSF1R also signals to a wide range of transcriptional factors, including MYC and FOS¹⁵. MYC signalling has been shown to be important for pro-tumoral phenotypes¹⁰⁰. CSF1R expression is regulated in turn by ETS2 transcription factors, and genetic ablation of this factor in macrophages in PyMT tumours recapitulates the loss of CSF1 in tumours, as angiogenesis is inhibited and tumour growth decreases¹⁰¹. To study the interaction of these factors and other regulatory molecules such as microRNAs and epigenetic controls³ will require sophisticated genomic analyses that will help to differentiate the regulation of the multiple subsets²³. These functions and other regulatory systems have been reviewed recently³.

Macrophages in inflammatory disease

Macrophages have important roles in many chronic diseases, including atherosclerosis, asthma, inflammatory bowel disease, rheumatoid

arthritis and fibrosis^{7,102–104}. Their contributions to these diseases vary greatly in different stages of disease and are controlled by many factors. For example, allergic asthma is a complex chronic inflammatory disease of the lung defined by airway inflammation, airway obstruction, airway hyper-responsiveness and pathological lung remodelling. The inflammatory response is characterized by the recruitment of T_H2 lymphocytes, mast cells, eosinophils and macrophages to the lung, and by elevated expression of allergen-specific immunoglobulin-E (IgE) in the serum. It has been suggested that the chronicity of type 2 cytokine-mediated airway inflammation that is characteristic of allergic asthma is explained by the presence of a macrophage-like antigen-presenting cell population that persists in the airway lumen¹⁰⁵. Pulmonary macrophages produce a variety of factors that directly stimulate airway smooth-muscle contractility and degradation of the ECM that contributes to pathological airway remodelling. Airway macrophages from some asthmatics are bathed in type-2-associated cytokines, including IL-4, IL-13 and IL-33, causing their differentiation, which has been implicated in the pathogenesis of asthma². These macrophages in turn promote the production of type 2 cytokines by pulmonary CD4 T lymphocytes, and produce a variety of cytokines and chemokines that regulate the recruitment of eosinophils, T_H2 cells and basophils to the lung, suggesting a viscous cycle that worsens disease⁷. Adoptive transfer studies have shown that the severity of allergen-induced disease is exacerbated by IL-4R⁺ macrophages¹⁰⁶, whereas protection from allergic airway disease is associated with a reduction in IL-4R⁺ macrophages in some studies¹⁰⁷. Increased numbers of IL-4R⁺ macrophages have also been reported in the lungs of asthmatic patients that have reduced lung function¹⁰⁸. Nevertheless, studies conducted with LysM^{Cre} IL-4R α ^{-lox} mice in which Cre-mediated recombination results in deletion of the IL-4R α chain in the myeloid cell lineage identified no substantial role for IL-4R α -activated macrophages in ovalbumin- and house-dust-mite-induced allergic airway disease¹⁰⁹.

Macrophages have also been implicated in the pathogenesis of a variety of autoimmune disease, including rheumatoid arthritis, multiple sclerosis and inflammatory bowel diseases. In these diseases, macrophages are an important source of many of the key inflammatory cytokines that have been identified as drivers of autoimmune inflammation, including IL-12, IL-18, IL-23 and TNF- α ¹⁰. Macrophage-derived IL-23 promotes end-stage joint autoimmune inflammation in mice. TNF- α also functions as an important driver of chronic polyarthritis, whereas IFN- γ - and TNF- α -dependent arthritis in mice has been attributed to macrophages and dendritic cells that produce IL-18 and IL-12. The pathogenesis of chronic demyelinating diseases of the central nervous system (CNS) has also been attributed to macrophages that display a pro-inflammatory phenotype. These inflammatory macrophages contribute to axon demyelination in experimental autoimmune encephalomyelitis in mice, a frequently used model of multiple sclerosis. Consequently, novel therapeutic strategies that target specific myeloid cell populations could help to ameliorate pathogenic inflammation in the CNS¹¹¹. The pathogenesis of inflammatory bowel disease is also tightly regulated by inflammatory macrophages. A subset of TLR2⁺CCR2⁺CX3CR1^{int} Ly6c^{hi}GR1⁺ macrophages has been shown to promote colonic inflammation by producing TNF- α ¹¹². A recent study showed that inflammatory mediators produced in the colon convert homeostatic anti-inflammatory macrophages into pro-inflammatory dendritic-cell-like cells that are capable of producing large quantities of IL-12, IL-23, inducible nitric oxides synthase and TNF- α ¹¹³. CD14⁺ macrophages that produce IL-23 and TNF- α have also been identified in Crohn's disease patients¹⁰³. Thus, macrophages and dendritic cells are key producers of many of the cytokines that have been implicated in the pathogenesis of inflammatory bowel disease.

Although there is substantial evidence to support the idea that inflammatory macrophages have roles in autoimmune inflammation, many studies have also reported suppressive roles for macrophages. For example, macrophages that produce reactive oxygen species can protect mice from arthritis by inhibiting T-cell activation¹¹⁴. Pro-inflammatory cytokines that are produced by activated macrophages have also been

shown to protect mice from Crohn's disease by facilitating the clearance of pathogenic commensal bacteria from the mucosal lining of the bowel¹¹⁵. Recruited monocytes and resident tissue macrophages are also thought to maintain homeostasis in the intestine by clearing apoptotic cells and debris, promoting epithelial repair, antagonizing pro-inflammatory macrophages, and by producing the suppressive cytokine IL-10, which is critical for the maintenance of FOXP3 expression in colonic regulatory T cells (T_{reg} cells)^{113,115,116}. Macrophages also protect rodents from demyelinating diseases of the CNS by promoting T-cell apoptosis and by expressing the anti-inflammatory cytokines TGF- β 1 and IL-10. The inhibitory receptor CD200 (also known as OX2), which is also expressed on anti-inflammatory macrophages, has been shown to prevent the onset of experimental autoimmune encephalomyelitis in mice¹¹⁷. A unique population of monocyte-derived macrophages also reduces inflammation resulting from spinal cord injury, providing further evidence of a protective role for macrophages in the CNS¹¹⁸. Together, these observations show how changes in macrophage differentiation in the local environment can have a decisive role in the pathogenesis of a wide variety of autoimmune and inflammatory diseases.

Macrophages in fibrosis

Although macrophages phagocytose and clear apoptotic cells as a part of their normal homeostatic function in tissues, when they encounter invading organisms or necrotic debris after injury, they become activated by endogenous danger signals and pathogen-associated molecular patterns. These activated macrophages produce anti-microbial mediators, like reactive oxygen and nitrogen species and proteinases, that help to kill invading pathogens and thus assist in the restoration of tissue homeostasis. However, they also produce a variety of inflammatory cytokines and chemokines such as TNF- α , IL-1, IL-6 and CCL2 that help to drive inflammatory and anti-microbial responses forward^{8,72}. This exacerbates tissue injury and in some cases leads to aberrant wound healing and ultimately fibrosis (scarring) if the response is not adequately controlled, as has been demonstrated by the selective depletion of macrophages at various stages of the wound-healing response¹¹⁹. Therefore, in recent years research has focused on elucidating the mechanisms that suppress inflammation and prevent the development of fibrosis. Although most wound-healing responses are self-limiting once the tissue-damaging irritant is removed, in many chronic fibrotic diseases the irritant is either unknown or cannot be eliminated easily¹²⁰. In this situation, it is crucial that the dominant macrophage population converts from one exhibiting a pro-inflammatory phenotype to one exhibiting anti-inflammatory, suppressive or regulatory characteristics so that collateral tissue damage is kept at a minimum (Fig. 4). A variety of mediators and mechanisms have been shown to regulate this conversion, including the cytokines IL-4 and IL-13, Fc γ receptor and TLR signalling, the purine nucleoside adenosine and A2A receptor signalling, prostaglandins, T_{reg} cells, and B1 B cells^{120,121}. Each of these mediators has been shown to activate distinct populations of macrophages with suppressive or regulatory characteristics. These 'regulatory' macrophages express a variety of soluble mediators, signalling intermediates and cell-surface receptors, including IL-10, arginase 1, IKK α , MMP13, maresins, CD200, RELM α and PD-L2, which have all been shown to decrease the magnitude and/or duration of inflammatory responses, and in some cases to contribute to the resolution of fibrosis⁷. They also produce a variety of soluble mediators, including CSF1, insulin-like growth factor 1, and VEGF, that promote wound healing¹²². Consequently, in addition to promoting fibrosis, macrophages are intimately involved in the recovery phase of fibrosis by inducing ECM degradation, phagocytizing apoptotic myofibroblasts and cellular debris, and by dampening the immune response that contributes to tissue injury¹²⁰. Therefore, current fibrosis research is focused on characterizing these regulatory macrophage populations and devising therapeutic strategies that can exploit their anti-inflammatory, anti-fibrotic and wound-healing properties.

Perspectives

Our understanding of macrophage biology is increasing rapidly, and it is now understood that they have diverse origins, transcriptional complexity and lability, and are capable of phenotypic switching in accordance with homeostatic demands and in response to insult. Macrophages are involved in almost every disease and represent attractive therapeutic targets because their function can be augmented or inhibited to alter disease outcome. However, for these therapies to be effective it is necessary to understand macrophage diversity and define their phenotypes according to anatomical location and function, and according to the regulation of the particular set-points that define the recognizable macrophages, such as microglia, osteoclasts and Kupffer cells. Indeed, the recognition of multiple origins (yolk sac, fetal liver, bone marrow) may result in the conclusion that there is no such thing as a 'macrophage' but instead, clades of cells that have similar characteristics but different origins. Their different origins may in fact provide unique opportunities to target the recruited monocytes and macrophages selectively in the context of the chronic diseases discussed above, thereby inhibiting the pathology without disturbing resident macrophages and thereby maintaining normal homeostasis. To define these similarities and differences it will be necessary to determine proteomes and transcriptomes of particular subtypes; this was recently performed for resident macrophages¹. The field of genomic analysis is advancing rapidly and will provide unique insights and novel methods to define macrophage types. Furthermore, macrophage biology in humans is poorly developed because of the technical limitations of obtaining fresh material for fluorescence-associated cell sorting (FACS) and the over-reliance of functional and genomic studies on cell lines such as the myelomonocytic leukaemic cell line THP1 (ref. 123) or the *in vitro* differentiation of circulating monocytes by CSF1. Notable differences also exist between human and mouse macrophages; for example, the inability of human macrophages to increase arginase 1 expression that is an important marker of IL-4-regulated macrophages in mice³. These differences mean that the binary classifications such as M1 and M2 are inadequate. Human macrophage diversity has begun to be defined¹²⁴; several sequencing efforts are in progress and these will begin to address the essential need to translate mouse biology into the human context.

Considerable advances in our knowledge of macrophage biology have been made recently using mouse genetic approaches. For example, macrophages can be fluorescently labelled by expressing GFP from the *Csf1r* promoter, and this is used to identify and, in some cases, record live images of them using intravital microscopy^{23,125}. Furthermore, the development of macrophage-restricted Cre recombinases—for example, expressed from the *LysM* or *Csf1r* promoters—and the ability to ablate macrophages through the expression of the diphtheria toxin receptor, which sensitizes mouse cells to the toxic effect of diphtheria toxin¹¹⁹, or using miRNAs to direct the expression of herpes simplex virus thymidine kinase in macrophages, have been key to defining the functions of macrophages. Although these systems have provided notable insights into macrophage function, none of the promoters is uniquely expressed in macrophages, and they are also expressed in most macrophage types, thereby making it difficult to discriminate the functions of subclasses of macrophages. In the future, specific promoters will be developed to ablate genes in particular subsets, more sophisticated lineage tracing will make it possible to follow cell fates, and subtype switching will be possible through photo-activatable fluors such as Dendra2 that enable a single cell, or a few cells, to be tracked¹²⁵.

Therapeutic targeting of macrophages is now in progress^{21,23}. Most of the therapies are targeted at pan-macrophage markers such as CSF1R. In the case of CSF1R reagents, including small molecules and monoclonal antibodies that inhibit the ligand, ligand binding or tyrosine kinase activity of the receptor are at various stages of clinical trials for the treatment of cancer²¹. Other strategies in fibrosis and cancer have been to target the recruitment of macrophages, particularly through inhibition of inflammatory monocyte trafficking with anti-CCL2 or -CCR2 antibodies. In one example, the protective effects of recombinant human

serum amyloid P (also known as pentraxin 2) in idiopathic pulmonary fibrosis and post-surgical scarring in patients treated for glaucoma are thought to occur through the reduction of inflammation and fibrosis resulting from the induction of IL-10 production in regulatory macrophages¹⁰⁷. Neutralization of GM-CSF using antibodies is being tested in phase II trials for multiple sclerosis and rheumatoid arthritis²¹. In the future, it seems that it will be possible to exploit the inherent plasticity of macrophages to adjust their set points to control obesity by down-modulating inflammatory cytokines, to resolve fibrosis by inducing the differentiation of resolving macrophages, and to treat cancer by converting macrophages from their trophic to an immunologically activated anti-tumoral state.

Received 18 October 2012; accepted 20 February 2013.

1. Gautier, E. L. *et al.* Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nature Immunol.* **13**, 1118–1128 (2012).
This paper provides a detailed analysis of the macrophage transcriptome. Several novel genes are identified that are distinctly and universally associated with mature tissue-resident macrophages, but the results also illustrate the extreme diversity of these cell types.
2. Gordon, S. Alternative activation of macrophages. *Nature Rev. Immunol.* **3**, 23–35 (2003).
3. Sica, A. & Mantovani, A. Macrophage plasticity and polarization: *in vivo* veritas. *J. Clin. Invest.* **122**, 787–795 (2012).
4. Geissmann, F. *et al.* Development of monocytes, macrophages, and dendritic cells. *Science* **327**, 656–661 (2010).
5. Jenkins, S. J. *et al.* Local macrophage proliferation, rather than recruitment from the blood, is a signature of T_H2 inflammation. *Science* **332**, 1284–1288 (2011).
This paper shows that tissue macrophages can proliferate in response to IL-4, suggesting that monocyte recruitment and definitive haematopoiesis may not be required for macrophage expansion in type 2 immunity.
6. Schulz, C. *et al.* A lineage of myeloid cells independent of Myb and hematopoietic stem cells. *Science* **336**, 86–90 (2012).
Together with refs 10 and 11, this paper indicates that the mononuclear phagocytic lineage needs to be reassessed and that most resident adult macrophage populations derive from the yolk sac.
7. Murray, P. J. & Wynn, T. A. Protective and pathogenic functions of macrophage subsets. *Nature Rev. Immunol.* **11**, 723–737 (2011).
8. Wynn, T. A. & Barron, L. Macrophages: master regulators of inflammation and fibrosis. *Semin. Liver Dis.* **30**, 245–257 (2010).
A comprehensive review examining the regulatory role of macrophages in chronic inflammatory disease and fibrosis.
9. Gordon, S. & Taylor, P. R. Monocyte and macrophage heterogeneity. *Nature Rev. Immunol.* **5**, 953–964 (2005).
The definitive review of activated and alternatively activated macrophages, with detailed explanations of the definitions and restrictions of these terms.
10. Ginhoux, F. *et al.* Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science* **330**, 841–845 (2010).
11. Hoeffel, G. *et al.* Adult Langerhans cells derive predominantly from embryonic fetal liver monocytes with a minor contribution of yolk sac-derived macrophages. *J. Exp. Med.* **209**, 1167–1181 (2012).
12. Frankenberger, M. *et al.* A defect of CD16-positive monocytes can occur without disease. *Immunobiology* **218**, 169–174 (2013).
13. Hume, D. A. Macrophages as APC and the dendritic cell myth. *J. Immunol.* **181**, 5829–5835 (2008).
14. Satpathy, A. T., Wu, X., Albring, J. C. & Murphy, K. M. Re(de)fining the dendritic cell lineage. *Nature Immunol.* **13**, 1145–1154 (2012).
15. Chitu, V. & Stanley, E. R. Colony-stimulating factor-1 in immunity and inflammation. *Curr. Opin. Immunol.* **18**, 39–48 (2006).
16. Erblach, B., Zhu, L., Etgen, A. M., Dobrenis, K. & Pollard, J. W. Absence of colony stimulation factor-1 receptor results in loss of microglia, disrupted brain development and olfactory deficits. *PLoS ONE* **6**, e26317 (2011).
17. Wei, S. *et al.* Functional overlap but differential expression of CSF-1 and IL-34 in their CSF-1 receptor-mediated regulation of myeloid cells. *J. Leukoc. Biol.* **8**, 495–505 (2010).
18. Wang, Y. *et al.* IL-34 is a tissue-restricted ligand of CSF1R required for the development of Langerhans cells and microglia. *Nature Immunol.* **13**, 753–760 (2012).
19. Pollard, J. W. Trophic macrophages in development and disease. *Nature Rev. Immunol.* **9**, 259–270 (2009).
20. Niida, S. *et al.* Vascular endothelial growth factor can substitute for macrophage colony-stimulating factor in the support of osteoclastic bone resorption. *J. Exp. Med.* **190**, 293–298 (1999).
21. Hamilton, J. A. & Achuthan, A. Colony stimulating factors and myeloid cell biology in health and disease. *Trends Immunol.* **34**, 81–89 (2013).
22. Miller, J. C. *et al.* Deciphering the transcriptional network of the dendritic cell lineage. *Nature Immunol.* **13**, 888–899 (2012).
23. Hume, D. A. The complexity of constitutive and inducible gene expression in mononuclear phagocytes. *J. Leukoc. Biol.* (2012).
24. Edwards, J. R. & Mundy, G. R. Advances in osteoclast biology: old findings and new insights from mouse models. *Nature Rev. Rheumatol.* **7**, 235–243 (2011).
25. Bigley, V. *et al.* The human syndrome of dendritic cell, monocyte, B and NK lymphoid deficiency. *J. Exp. Med.* **208**, 227–234 (2011).
26. Stefater, J. A. III, Ren, S., Lang, R. A. & Duffield, J. S. Metchnikoff's policeman: macrophages in development, homeostasis and regeneration. *Trends Mol. Med.* (2011).
27. Gyorki, D. E., Asselin-Labat, M. L., van Rooijen, N., Lindeman, G. J. & Visvader, J. E. Resident macrophages influence stem cell activity in the mammary gland. *Breast Cancer Res.* **11**, R62 (2009).
28. Pull, S. L., Doherty, J. M., Mills, J. C., Gordon, J. I. & Stappenbeck, T. S. Activated macrophages are an adaptive element of the colonic epithelial progenitor niche necessary for regenerative responses to injury. *Proc. Natl Acad. Sci. USA* **102**, 99–104 (2005).
29. Chow, A. *et al.* Bone marrow CD169⁺ macrophages promote the retention of hematopoietic stem and progenitor cells in the mesenchymal stem cell niche. *J. Exp. Med.* **208**, 261–271 (2011).
This paper, together with refs 27, 28 and 30, shows that macrophages regulate various stem cell niches.
30. Boulter, L. *et al.* Macrophage-derived Wnt opposes Notch signaling to specify hepatic progenitor cell fate in chronic liver disease. *Nature Med.* **18**, 572–579 (2012).
31. Kawane, K. *et al.* Requirement of DNase II for definitive erythropoiesis in the mouse fetal liver. *Science* **292**, 1546–1549 (2001).
32. Jaiswal, S. *et al.* CD47 is upregulated on circulating hematopoietic stem cells and leukemia cells to avoid phagocytosis. *Cell* **138**, 271–285 (2009).
33. Gordy, C., Pua, H., Sempowski, G. D. & He, Y. W. Regulation of steady-state neutrophil homeostasis by macrophages. *Blood* **117**, 618–629 (2011).
34. Dai, X. M. *et al.* Targeted disruption of the mouse CSF-1 receptor gene results in osteopetrosis, mononuclear phagocyte deficiency, increased primitive progenitor cell frequencies and reproductive defects. *Blood* **99**, 111–120 (2002).
35. Savill, J., Dransfield, I., Gregory, C. & Haslett, C. A blast from the past: clearance of apoptotic cells regulates immune responses. *Nature Rev. Immunol.* **2**, 965–975 (2002).
36. Rao, S. *et al.* Obligatory participation of macrophages in an angiopoietin 2-mediated cell death switch. *Development* **134**, 4449–4458 (2007).
37. Stefater, J. A. III *et al.* Regulation of angiogenesis by a non-canonical Wnt-Fit1 pathway in myeloid cells. *Nature* **474**, 511–515 (2011).
An important paper showing the molecular basis of the macrophage regulation of angiogenesis through the WNT pathway.
38. Tammela, T. *et al.* VEGFR-3 controls tip to stalk conversion at vessel fusion sites by reinforcing Notch signalling. *Nature Cell Biol.* **13**, 1202–1213 (2011).
39. Fantin, A. *et al.* Tissue macrophages act as cellular chaperones for vascular anastomosis downstream of VEGF-mediated endothelial tip cell induction. *Blood* **116**, 829–840 (2010).
40. Gordon, E. J. *et al.* Macrophages define dermal lymphatic vessel calibre during development by regulating lymphatic endothelial cell proliferation. *Development* **137**, 3899–3910 (2010).
41. Machnik, A. *et al.* Macrophages regulate salt-dependent volume and blood pressure by a vascular endothelial growth factor-C-dependent buffering mechanism. *Nature Med.* **15**, 545–552 (2009).
42. Nandi, S. *et al.* The CSF-1 receptor ligands IL-34 and CSF-1 exhibit distinct developmental brain expression patterns and regulate neural progenitor cell maintenance and maturation. *Dev. Biol.* **367**, 100–113 (2012).
43. Li, Y., Du, X. F., Liu, C. S., Wen, Z. L. & Du, J. L. Reciprocal regulation between resting microglial dynamics and neuronal activity *in vivo*. *Dev. Cell* **23**, 1189–1202 (2012).
A paper that demonstrates that microglia regulate neuronal activity in zebrafish using intravital imaging.
44. Paolicelli, R. C. *et al.* Synaptic pruning by microglia is necessary for normal brain development. *Science* **333**, 1456–1458 (2011).
45. Rademakers, R. *et al.* Mutations in the colony stimulating factor 1 receptor (CSF1R) gene cause hereditary diffuse leukoencephalopathy with spheroids. *Nature Genet.* **44**, 200–205 (2011).
46. London, A. *et al.* Neuroprotection and progenitor cell renewal in the injured adult murine retina requires healing monocyte-derived macrophages. *J. Exp. Med.* **208**, 23–39 (2011).
47. Kigerl, K. A. *et al.* Identification of two distinct macrophage subsets with divergent effects causing either neurotoxicity or regeneration in the injured mouse spinal cord. *J. Neurosci.* **29**, 13435–13444 (2009).
48. Salegio, E. A., Pollard, A. N., Smith, M. & Zhou, X. F. Macrophage presence is essential for the regeneration of ascending afferent fibres following a conditioning sciatic nerve lesion in adult rats. *BMC Neurosci.* **12**, 11 (2011).
49. Hotamisligil, G. S. Inflammation and metabolic disorders. *Nature* **444**, 860–867 (2006).
50. Chawla, A., Nguyen, K. D. & Goh, Y. P. Macrophage-mediated inflammation in metabolic disease. *Nature Rev. Immunol.* **11**, 738–749 (2011).
51. Odegaard, J. I. & Chawla, A. Pleiotropic actions of insulin resistance and inflammation in metabolic homeostasis. *Science* **339**, 172–177 (2013).
52. Olefsky, J. & Glass, C. Macrophages, inflammation, and insulin resistance. *Annu. Rev. Physiol.* **72**, 219–246 (2010).
53. Rosen, E. D. & Spiegelman, B. M. Adipocytes as regulators of energy balance and glucose homeostasis. *Nature* **444**, 847–853 (2006).
54. Lumeng, C. N., Bodzin, J. L. & Saltiel, A. R. Obesity induces a phenotypic switch in adipose tissue macrophage polarization. *J. Clin. Invest.* **117**, 175–184 (2007).
55. Weisberg, S. P. *et al.* Obesity is associated with macrophage accumulation in adipose tissue. *J. Clin. Invest.* **112**, 1796–1808 (2003).

56. Xu, H. *et al.* Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J. Clin. Invest.* **112**, 1821–1830 (2003).
This paper, together with ref. 55, was the first to demonstrate that obesity results in infiltration of WAT by macrophages, which contributes to its inflamed nature.
57. Sun, K., Kusminski, C. M. & Scherer, P. E. Adipose tissue remodeling and obesity. *J. Clin. Invest.* **121**, 2094–2101 (2011).
58. Kang, K. *et al.* Adipocyte-derived Th2 cytokines and myeloid PPAR δ regulate macrophage polarization and insulin sensitivity. *Cell Metab.* **7**, 485–495 (2008).
59. Liao, X. *et al.* 7, 485–495 Kruppel-like factor 4 regulates macrophage polarization. *J. Clin. Invest.* **121**, 2736–2749 (2011).
60. Odegaard, J. I. *et al.* Macrophage-specific PPAR γ controls alternative activation and improves insulin resistance. *Nature* **447**, 1116–1120 (2007).
This paper showed that residence of AAMs in WAT is necessary for the maintenance of insulin sensitivity in obese animals.
61. Odegaard, J. I. *et al.* Alternative M2 activation of Kupffer cells by PPAR δ ameliorates obesity-induced insulin resistance. *Cell Metab.* **7**, 496–507 (2008).
62. Wu, D. *et al.* Eosinophils sustain adipose alternatively activated macrophages associated with glucose homeostasis. *Science* **332**, 243–247 (2011).
63. Weisberg, S. P. *et al.* CCR2 modulates inflammatory and metabolic effects of high-fat feeding. *J. Clin. Invest.* **116**, 115–124 (2006).
64. Cinti, S. *et al.* Adipocyte death defines macrophage localization and function in adipose tissue of obese mice and humans. *J. Lipid Res.* **46**, 2347–2355 (2005).
65. Lowell, B. B. & Spiegelman, B. M. Towards a molecular understanding of adaptive thermogenesis. *Nature* **404**, 652–660 (2000).
66. Nguyen, K. D. *et al.* Alternatively activated macrophages produce catecholamines to sustain adaptive thermogenesis. *Nature* **480**, 104–108 (2011).
This paper demonstrated a physiological function for AAMs in sustaining adaptive thermogenesis, which allows mammals to adapt to cold environments.
67. Samuel, V. T. & Shulman, G. I. Mechanisms for insulin resistance: common threads and missing links. *Cell* **148**, 852–871 (2012).
68. Huang, W. *et al.* Depletion of liver Kupffer cells prevents the development of diet-induced hepatic steatosis and insulin resistance. *Diabetes* **59**, 347–357 (2010).
69. Eguchi, K. *et al.* Saturated fatty acid and TLR signaling link beta cell dysfunction and islet inflammation. *Cell Metab.* **15**, 518–533 (2012).
70. Ehses, J. A. *et al.* Increased number of islet-associated macrophages in type 2 diabetes. *Diabetes* **56**, 2356–2370 (2007).
71. Nathan, C. & Ding, A. Nonresolving inflammation. *Cell* **140**, 871–882 (2010).
72. Sindrilaru, A. *et al.* An unrestrained proinflammatory M1 macrophage population induced by iron impairs wound healing in humans and mice. *J. Clin. Invest.* **121**, 985–997 (2011).
73. Krausgruber, T. *et al.* IRF5 promotes inflammatory macrophage polarization and T_H1–T_H17 responses. *Nature Immunol.* **12**, 231–238 (2011).
This paper showed that IRF5 expression is induced in macrophages in response to inflammatory stimuli and that this contributes to the polarization of macrophages with an inflammatory phenotype, which causes T_H1 and T_H17 cells to respond.
74. Ahn, G. O. *et al.* Inhibition of Mac-1 (CD11b/CD18) enhances tumor response to radiation by reducing myeloid cell recruitment. *Proc. Natl Acad. Sci. USA* **107**, 8363–8368 (2010).
75. Qian, B. Z. & Pollard, J. W. Macrophage diversity enhances tumor progression and metastasis. *Cell* **141**, 39–51 (2010).
76. Balkwill, F. R. & Mantovani, A. Cancer-related inflammation: common themes and therapeutic opportunities. *Semin. Cancer Biol.* **22**, 33–40 (2012).
77. Zaidi, M. R. *et al.* Interferon-gamma links ultraviolet radiation to melanomagenesis in mice. *Nature* **469**, 548–553 (2011).
78. Deng, L. *et al.* A novel mouse model of inflammatory bowel disease links mammalian target of rapamycin-dependent hyperproliferation of colonic epithelium to inflammation-associated tumorigenesis. *Am. J. Pathol.* **176**, 952–967 (2010).
79. Arthur, J. C. *et al.* Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* **338**, 120–123 (2012).
80. DeNardo, D. G., Andreu, P. & Coussens, L. M. Interactions between lymphocytes and myeloid cells regulate pro- versus anti-tumor immunity. *Cancer Metastasis Rev.* **29**, 309–316 (2010).
81. Hanahan, D. & Coussens, L. M. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* **21**, 309–322 (2012).
82. Condeelis, J. & Pollard, J. W. Macrophages: obligate partners for tumor cell migration, invasion, and metastasis. *Cell* **124**, 263–266 (2006).
83. Chen, J. *et al.* CCL18 from tumor-associated macrophages promotes breast cancer metastasis via PITPNM3. *Cancer Cell* **19**, 541–555 (2011).
84. Mason, S. D. & Joyce, J. A. Proteolytic networks in cancer. *Trends Cell Biol.* **21**, 228–237 (2011).
85. Kessenbrock, K., Plaks, V. & Werb, Z. Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell* **141**, 52–67 (2010).
86. Lin, E. Y. *et al.* Macrophages regulate the angiogenic switch in a mouse model of breast cancer. *Cancer Res.* **66**, 11238–11246 (2006).
87. Mazzei, R. *et al.* Targeting the ANG2/TIE2 axis inhibits tumor growth and metastasis by impairing angiogenesis and disabling rebounds of proangiogenic myeloid cells. *Cancer Cell* **19**, 512–526 (2011).
88. Murdoch, C., Muthana, M., Coffelt, S. B. & Lewis, C. E. The role of myeloid cells in the promotion of tumour angiogenesis. *Nature Rev. Cancer* **8**, 618–631 (2008).
89. Stockmann, C. *et al.* Deletion of vascular endothelial growth factor in myeloid cells accelerates tumorigenesis. *Nature* **456**, 814–818 (2008).
90. Shojai, F. *et al.* Tumor refractoriness to anti-VEGF treatment is mediated by CD11b⁺Gr1⁺ myeloid cells. *Nature Biotechnol.* **25**, 911–920 (2007).
91. Du, R. *et al.* HIF1 α induces the recruitment of bone marrow-derived vascular modulatory cells to regulate tumor angiogenesis and invasion. *Cancer Cell* **13**, 206–220 (2008).
92. Lin, E. Y. *et al.* Vascular endothelial growth factor restores delayed tumor progression in tumors depleted of macrophages. *Mol. Oncol.* **1**, 288–302 (2007).
93. Psaila, B. & Lyden, D. The metastatic niche: adapting the foreign soil. *Nature Rev. Cancer* **9**, 285–293 (2009).
94. Peinado, H. *et al.* Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. *Nature Med.* **18**, 883–891 (2012).
95. Gil-Bernabé, A. M. *et al.* Recruitment of monocytes/macrophages by tissue factor-mediated coagulation is essential for metastatic cell survival and premetastatic niche establishment in mice. *Blood* **119**, 3164–3175 (2012).
96. Qian, B. Z. *et al.* CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature* **475**, 222–225 (2011).
This paper suggests that macrophages may represent a therapeutic target to prevent tumour cell metastatic seeding and growth.
97. Coussens, L. M. & Pollard, J. W. Leukocytes in mammary development and cancer. *Cold Spring Harb. Perspect. Biol.* **3**, a003285 (2011).
98. Wu, Y. & Zheng, L. Dynamic education of macrophages in different areas of human tumors. *Cancer Microenvironment* **5**, 195–201 (2012).
99. Hagemann, T. *et al.* “Re-educating” tumor-associated macrophages by targeting NF- κ B. *J. Exp. Med.* **205**, 1261–1268 (2008).
100. Pello, O. M. *et al.* Role of c-MYC in alternative activation of human macrophages and tumor-associated macrophage biology. *Blood* **119**, 411–421 (2012).
101. Zabuawala, T. *et al.* An ets2-driven transcriptional program in tumor-associated macrophages promotes tumor metastasis. *Cancer Res.* **70**, 1323–1333 (2010).
102. Hansson, G. K. & Hermansson, A. The immune system in atherosclerosis. *Nature Immunol.* **12**, 204–212 (2011).
103. Kamada, N. *et al.* Unique CD14 intestinal macrophages contribute to the pathogenesis of Crohn disease via IL-23/IFN- γ axis. *J. Clin. Invest.* **118**, 2269–2280 (2008).
104. Libby, P., Ridker, P. M. & Hansson, G. K. Progress and challenges in translating the biology of atherosclerosis. *Nature* **473**, 317–325 (2011).
105. Julia, V. *et al.* A restricted subset of dendritic cells captures airborne antigens and remains able to activate specific T cells long after antigen exposure. *Immunity* **16**, 271–283 (2002).
106. Ford, A. Q. *et al.* Adoptive transfer of IL-4R α ⁺ macrophages is sufficient to enhance eosinophilic inflammation in a mouse model of allergic lung inflammation. *BMC Immunol.* **13**, 6 (2012).
107. Moreira, A. P. *et al.* Serum amyloid P attenuates M2 macrophage activation and protects against fungal spore-induced allergic airway disease. *J. Allergy Clin. Immunol.* **126**, 712–721 (2010).
108. Melgert, B. N. *et al.* More alternative activation of macrophages in lungs of asthmatic patients. *J. Allergy Clin. Immunol.* **127**, 831–833 (2011).
109. Nieuwenhuizen, N. E. *et al.* Allergic airway disease is unaffected by the absence of IL-4R α -dependent alternatively activated macrophages. *J. Allergy Clin. Immunol.* **130**, 743–750 (2012).
110. Murray, P. J. & Wynn, T. A. Obstacles and opportunities for understanding macrophage polarization. *J. Leukoc. Biol.* **89**, 557–563 (2011).
111. Ponomarev, E. D., Veremeyko, T., Barteneva, N., Krichevsky, A. M. & Weiner, H. L. MicroRNA-124 promotes microglia quiescence and suppresses EAE by deactivating macrophages via the C/EBP- α -PU.1 pathway. *Nature Med.* **17**, 64–70 (2011).
112. Platt, A. M., Bain, C. C., Bordon, Y., Sester, D. P. & Mowat, A. M. An independent subset of TLR expressing CCR2-dependent macrophages promotes colonic inflammation. *J. Immunol.* **184**, 6843–6854 (2010).
113. Rivollier, A., He, J., Kole, A., Valatas, V. & Kelsall, B. L. Inflammation switches the differentiation program of Ly6C^{hi} monocytes from antiinflammatory macrophages to inflammatory dendritic cells in the colon. *J. Exp. Med.* **209**, 139–155 (2012).
114. Gelderman, K. A. *et al.* Macrophages suppress T cell responses and arthritis development in mice by producing reactive oxygen species. *J. Clin. Invest.* **117**, 3020–3028 (2007).
115. Smith, A. M. *et al.* Disordered macrophage cytokine secretion underlies impaired acute inflammation and bacterial clearance in Crohn’s disease. *J. Exp. Med.* **206**, 1883–1897 (2009).
116. Murai, M. *et al.* Interleukin 10 acts on regulatory T cells to maintain expression of the transcription factor Foxp3 and suppressive function in mice with colitis. *Nature Immunol.* **10**, 1178–1184 (2009).
117. Hoek, R. M. *et al.* Down-regulation of the macrophage lineage through interaction with OX2 (CD200). *Science* **290**, 1768–1771 (2000).
118. Shechter, R. *et al.* Infiltrating blood-derived macrophages are vital cells playing an anti-inflammatory role in recovery from spinal cord injury in mice. *PLoS Med.* **6**, e1000113 (2009).
119. Duffield, J. S. *et al.* Selective depletion of macrophages reveals distinct, opposing roles during liver injury and repair. *J. Clin. Invest.* **115**, 56–65 (2005).
120. Wynn, T. A. & Ramalingam, T. R. Mechanisms of fibrosis: therapeutic translation for fibrotic disease. *Nature Med.* **18**, 1028–1040 (2012).
121. Mosser, D. M. & Edwards, J. P. Exploring the full spectrum of macrophage activation. *Nature Rev. Immunol.* **8**, 958–969 (2008).

122. Thomas, J. A. *et al.* Macrophage therapy for murine liver fibrosis recruits host effector cells improving fibrosis, regeneration, and function. *Hepatology* **53**, 2003–2015 (2011).
123. Ravasi, T. *et al.* An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* **140**, 744–752 (2010).
124. Frankenberger, M. *et al.* Transcript profiling of CD16-positive monocytes reveals a unique molecular fingerprint. *Eur. J. Immunol.* **42**, 957–974 (2012).
125. Entenberg, D. *et al.* Setup and use of a two-laser multiphoton microscope for multi-channel intravital fluorescence imaging. *Nature Protocols* **6**, 1500–1520 (2011).

Acknowledgements The authors apologize to colleagues whose papers they were unable to cite on this occasion. T.A.W. is supported by the intramural research program of the National Institutes of Allergy and Infectious Diseases (National Institutes of Health (NIH)). This work was supported by the National Cancer Institute of the NIH

(award numbers R01CA131270S, U54HD058155 and P01CA100324 (to J.W.P.), and HL076746, DK094641 and DK094641 (to A.C.)); the Diabetes Family Fund (to the University of California, San Francisco), an American Heart Association (AHA) Innovative Award (12PILT11840038) and a NIH Director's Pioneer Award (DP1AR064158 to A.C.).

Author Contributions T.A.W., A.C. and J.W.P. contributed to the writing and editing of all aspects of this Review.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.A.W. (twynn@niaid.nih.gov), A.C. (ajay.chawla@ucsf.edu) or J.W.P. (jeff.pollard@ed.ac.uk).

Classical command of quantum systems

Ben W. Reichardt¹, Falk Unger² & Umesh Vazirani³

Quantum computation and cryptography both involve scenarios in which a user interacts with an imperfectly modelled or ‘untrusted’ system. It is therefore of fundamental and practical interest to devise tests that reveal whether the system is behaving as instructed. In 1969, Clauser, Horne, Shimony and Holt proposed an experimental test that can be passed by a quantum-mechanical system but not by a system restricted to classical physics. Here we extend this test to enable the characterization of a large quantum system. We describe a scheme that can be used to determine the initial state and to classically command the system to evolve according to desired dynamics. The bipartite system is treated as two black boxes, with no assumptions about their inner workings except that they obey quantum physics. The scheme works even if the system is explicitly designed to undermine it; any misbehaviour is detected. Among its applications, our scheme makes it possible to test whether a claimed quantum computer is truly quantum. It also advances towards a goal of quantum cryptography: namely, the use of ‘untrusted’ devices to establish a shared random key, with security based on the validity of quantum physics.

Do the laws of quantum mechanics place any limits on how well a classical experimentalist can characterize the state and dynamics of a large quantum system? As a thought experiment, consider that we are presented with a quantum system, together with instructions on how to control its evolution from a claimed initial state. Our goal is to determine if the system was indeed initialized as claimed, and if its state evolves as instructed.

More formally, we model the quantum system as a black box, with (for example) buttons and light bulbs to allow for classical interactions in binary. Using this limited interface, we wish to characterize the initial state of the system. We also wish to verify that on command—by pressing a suitable sequence of buttons—the system applies a chosen local Hamiltonian, or equivalently a sequence of local quantum gates, and outputs desired measurement results.

A positive answer to this fundamental question would have important consequences. First, as the power of quantum mechanics is harnessed at larger scales—with the advent of quantum computers—it will be useful to evaluate whether a quantum device in fact carries out the claimed dynamics^{1,2}. Second, the goal of quantum cryptography is to create cryptographic systems with security premised on basic laws of physics. Although this seemed to have been achieved with quantum key distribution (QKD) and its security proofs^{3–5}, attackers have repeatedly breached the security of QKD experiments by exploiting imperfect implementations of the quantum devices^{6–8}. Rather than relying on *ad hoc* countermeasures, Mayers and Yao’s vision⁹ of device-independent (DI) QKD, hinted at earlier in ref. 10, relaxes all modelling assumptions about the devices, and even allows for them to have been constructed by an adversary. It instead imagines giving the devices tests that cannot be passed unless they carry out the QKD protocol securely. The challenge at the heart of this vision is for an experimentalist to force untrusted quantum devices to act according to certain specifications. DIQKD has not been shown to be possible; the security proofs, first given in ref. 11, have required the unrealistic assumption that the devices have no memory between trials, or that each party has many, strictly isolated devices^{12–20}. A scheme for characterizing and commanding a black-box quantum device would provide a novel approach to achieving DIQKD.

The existence of a general scheme for commanding an adversarial quantum device appears singularly implausible. For example, in an adversarial setting, experiments cannot be repeated exactly to gather statistics, because a system with memory could deliberately deceive the experimentalist. More fundamentally, as macroscopic, classical entities, our access to a quantum system is extremely limited and indirect, and the measurements we apply collapse the quantum state. Furthermore, whereas the dimension of the underlying Hilbert space scales exponentially with the number of particles or can be infinite, the information accessible via measurement grows only linearly²¹. Indeed, as formulated it is impossible to command a single black-box system. Quite simply, it is impossible to distinguish between a quantum system that evolves as desired and a device that merely simulates the desired evolution using a classical computer.

In this Article, we consider a closely related scenario. Suppose we are instead given two devices, each modelled as a black box as above and prevented from communicating with the other. In this setting, with no further assumptions, we show how to command the devices classically. That is, there is a strategy for pushing the buttons such that the answering light bulb flashes will satisfy a prescribed test only if the two devices started in a particular initial quantum state, to which they applied a desired sequence of quantum gates. Moreover, the scheme is theoretically efficient, in the sense that the total effort, measured by the number of button pushes, scales as a polynomial function of the size of the desired quantum circuit. A DIQKD scheme follows, although it is far from practical.

Detailed overview

Rigidity of the CHSH test for quantumness

The starting point for our protocol is the famous Bell experiment²², and its subsequent ‘distillation’ by Clauser, Horne, Shimony and Holt²³ (CHSH). Conceptually modelled as a game (Fig. 1), it provides a test for ‘quantumness’, that is, a way for an experimentalist, whom we shall call Eve, to demonstrate the entanglement of two space-like separated devices, Alice and Bob. According to a Bell inequality, classical devices can win the game with probability at most 3/4. In contrast, quantum devices can win with probability $\omega^* = \cos^2(\pi/8) \approx 85.4\%$, which is optimal by Tsirelson’s inequality²⁴.

¹Electrical Engineering Department, University of Southern California, Los Angeles, California 90089, USA. ²Knight Capital Group, Inc., Santa Clara, California 95054, USA. ³Computer Science Division, University of California, Berkeley, California 94720, USA.

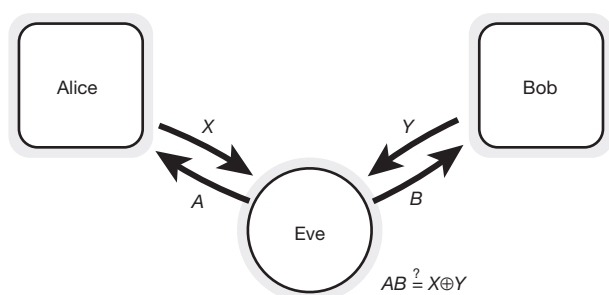


Figure 1 | Test for quantumness. In a CHSH experiment, or ‘game’, the experimentalist Eve sends random bits A and B to the devices Alice and Bob, respectively, who respond with bits X and Y . The devices ‘win’ if $AB = X \oplus Y$. Quantum devices can win with probability $\omega^* = \cos^2(\pi/8)$ if they follow an ideal CHSH strategy: on a shared EPR state $|\varphi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, Bob measures the Pauli operator σ_z if $B = 0$ or σ_x if $B = 1$, and Alice measures $(\sigma_z + (-1)^A \sigma_x)/\sqrt{2}$.

We prove a robust converse to Tsirelson’s inequality, namely a rigidity property of the CHSH game: nearly saturating Tsirelson’s bound locks into place the devices’ shared state and measurement operators. More precisely, if the devices win with probability $\omega^* - \varepsilon$, then they must share a state that is within a distance $O(\sqrt{\varepsilon})$ of an Einstein–Podolsky–Rosen (EPR) state, possibly in tensor product with an additional ancilla state. Moreover, their joint measurement strategy is necessarily $O(\sqrt{\varepsilon})$ -close to the ideal strategy from Fig. 1 (that is, applying Alice’s actual measurement operator to the shared state gets within distance $O(\sqrt{\varepsilon})$ of the result of applying her ideal measurement operator to the EPR state tensored with the ancilla; and similarly for Bob). Because each device can locate its qubit (quantum bit) share of the EPR state arbitrarily within its Hilbert space, these statements hold only up to local isometries.

A converse to Tsirelson’s inequality for the CHSH game has been shown previously in the exact case^{25,26}. Robustness is important for applications, however, because the success probability of a system can never be known exactly. A robust, $\varepsilon > 0$, converse statement has been shown for the game used in the original DIQKD proposal²⁷. Recently, robustness has independently been shown for the CHSH game^{28,29}.

Scalable test for quantumness

We scale up the CHSH test for quantumness to allow us to identify many qubits’ worth of entanglement. Consider a protocol in which Eve plays a long sequence of n CHSH games with Alice and Bob, and tests whether they win close to the optimal fraction ω^* of the games. Our main technical result, a multi-game rigidity theorem, establishes that if the devices pass Eve’s test with high probability, then at the beginning of a randomly chosen long subsequence of n^χ games, for some constant χ , Alice and Bob must share n^χ EPR states in tensor product, which they measure one at a time using the single-game ideal CHSH operators of Fig. 1. The j th game is played using the j th EPR state, different games being entirely independent. This is a step towards the general vision outlined above, because it characterizes the initial state of many qubits and allows Eve to command the devices to perform certain single-qubit operations. Of course, we cannot hope to characterize the devices’ strategies exactly, but only for a suitable notion of approximation.

The difficulty in proving this theorem is that although individual games are typically rigid, the states close to EPR states used in different games could overlap significantly. Furthermore, Alice and Bob’s strategy for playing each game—including, for example, the locations of the near EPR states—could depend on the previous games’ outcomes. The multi-game rigidity theorem rules out such wayward behaviour.

Verified quantum dynamics

The multi-game rigidity theorem gives strong control over the devices’ measurement operators for different games. As described

below, combining the CHSH game protocol with protocols for state and process tomography, and for computation by teleportation³⁰, gives a method for realizing arbitrary dynamics in quantum systems without making assumptions about their internal structure or operations. The dynamics are realized as the joint evolution of two isolated quantum systems, Alice and Bob, mediated by a classical experimentalist, Eve.

The problem of controlling computationally powerful but untrusted resources lies at the foundation of computer science. In the complexity class NP, for example, a polynomial-time routine—the ‘verifier’—is allowed one round of interaction with an arbitrarily powerful, but malicious, ‘prover’. We show that the same verifier can exploit the power of quantum mechanical provers³¹. In particular, (1) a classical verifier can efficiently simulate a quantum computer by interacting with two untrusted, polynomial-time quantum provers that share entanglement but cannot communicate between themselves. This delegated computation scheme is also ‘blind’, meaning that each prover learns nothing more about the computation than its length. Furthermore, (2) a classical verifier is as powerful as a quantum verifier in any interactions with multiple quantum provers (formally, the complexity classes QMIP and MIP* are equal).

Previous work introducing this problem has considered a ‘semi-quantum’ verifier, who manipulates a constant number of qubits while interacting with a prover^{1,2,32,33}. Our work is also inspired by a proposal³⁴ that QMIP should equal MIP*. Although our protocol also uses computation by teleportation, it has a very different form, based on the multi-game rigidity theorem.

Product structure from repeated games

A strategy \mathcal{S} for playing n sequential CHSH games specifies the initial joint state of Alice (A) and Bob (B) as well as their measurement operators for every possible situation. That is, for $D \in \{A, B\}$ and each $j = 1, \dots, n$, \mathcal{S} specifies the measurement operators used by device D in game (j, h_{j-1}^D) , where h_{j-1}^D is a transcript of the device’s input and output bits for the first $j-1$ games. A strategy \mathcal{S} induces a distribution on game transcripts. For two strategies to be ‘close’, the corresponding distributions on game transcripts should be close in total variation distance and, for almost all transcripts (drawn from either distribution), the resulting quantum states should be close in a suitable norm. We combine these conditions into one by defining for any strategy a block-diagonal density matrix that stores both the classical transcript and the resulting quantum state:

$$\rho_j = \bigoplus_{h_{j-1}} \Pr[h_{j-1}] \rho_j(h_{j-1}) \quad (1)$$

Here $h_{j-1} = (h_{j-1}^A, h_{j-1}^B)$ is the full transcript for the first $j-1$ games and $\rho_j(h_{j-1})$ is the state at the beginning of game j conditioned on h_{j-1} . Two strategies \mathcal{S} and $\tilde{\mathcal{S}}$ are close if the associated ρ_j and $\tilde{\rho}_j$ are close in trace distance ($\|\dots\|_{\text{tr}}$), for every j .

Assume that for every j and almost all h_{j-1} , the devices’ conditional joint strategy at the beginning of game j is ‘ ε -structured’, meaning that the devices win with probability at least $\omega^* - \varepsilon$. Our key theorem establishes that up to local basis changes, the devices’ initial state must be close to n EPR states, possibly in tensor product with an irrelevant extra state, and that their total strategy \mathcal{S} must be close to an ideal strategy $\hat{\mathcal{S}}$ that plays game j using the j th EPR state. Because the structure assumption can be established by standard statistical martingale arguments on $\text{poly}(n)$ sequential CHSH games, this implies the multi-game rigidity theorem.

The main challenge is to ‘locate’ the ideal strategy $\hat{\mathcal{S}}$ within Alice and Bob’s Hilbert space, that is, to find an isometry on each of their spaces under which their states and measurement operators are close to ideal. However, a priori, we do not know whether \mathcal{S} calls for the devices to measure actual qubits in each step, or, even if so, whether the qubits form EPR states, qubits for different games overlap each

other or the locations of the qubits depend on the outcomes of previous games.

A good place to start is the construction in the single-game rigidity theorem that locates the qubits. Consider an ε -structured strategy, consisting of some shared mixed state in $\mathcal{H}_A \otimes \mathcal{H}_B$, and two-outcome projective measurements for each of Eve's possible questions. Truncate the devices' Hilbert spaces to finitely many dimensions, then decompose each space by Jordan's lemma³⁵ into the direct sum of two-dimensional spaces invariant under the projections. Within each such two-dimensional subspace, adjust the projections so the angle between them matches that of the ideal strategy. This defines a $\{|0\rangle, |1\rangle\}$ basis for each subspace. Aligning the subspaces according to this basis allows each Hilbert space \mathcal{H}_D to be decomposed as the tensor product of a qubit and the remainder. See Supplementary Information and ref. 36 for proof details.

For multiple CHSH games, the given strategy \mathcal{S} can be transformed into a nearby ideal strategy $\tilde{\mathcal{S}}$ in a three-step sequence.

Step 1. Replace each device's measurement operators by the ideal operators known to exist from the CHSH rigidity theorem (for a single game). In the resulting strategy $\tilde{\mathcal{S}}$, each device D plays every game (j, h_{j-1}^D) using the ideal CHSH game operators on some qubit, up to a local change in basis. However, the basis change can depend arbitrarily on h_{j-1}^D , and the qubits for different values of j need not be in tensor product.

Step 2. In a 'multi-qubit ideal strategy', $\tilde{\mathcal{S}}$, the qubits used in each game can still depend on the local transcripts but must at least lie in tensor product with the qubits from previous games. This imposes a tensor-product subsystem structure that previous DIQKD proofs have assumed. The tensor-product structure is constructed beginning with a trivial transformation on $\tilde{\mathcal{S}}$: to each device, add n ancilla qubits each in state $|0\rangle$. Next, after a qubit has been measured, say as $|\alpha_j\rangle$ in game j , swap it with the j th ancilla qubit, and then rotate this fresh qubit from $|0\rangle$ to $|\alpha_j\rangle$ and continue playing games $j+1, \dots, n$. This defines a unitary change of basis that places the outcomes for games 1 to j in the first j ancilla qubits, and leaves the state in the original Hilbert space unchanged. At the end of the n games, undo the basis change: swap back the ancilla qubits and undo their rotations. Because qubits are set aside after being measured, the qubits for later games are automatically in tensor product with those for earlier games; the resulting strategy $\tilde{\mathcal{S}}$ is multi-qubit ideal.

Step 3. We replace $\tilde{\mathcal{S}}$ with an ideal strategy $\hat{\mathcal{S}}$, in which Alice and Bob each play using a fixed set of n qubits. Fix a transcript \hat{h}_n , chosen at random. For the first time, change the devices' initial state: replace ρ_1 with $\hat{\rho}_1$, a state having n EPR states in the locations determined by \hat{h}_n in $\tilde{\mathcal{S}}$. In $\hat{\mathcal{S}}$, the devices play using these EPR states, regardless of the actual transcript. This $\hat{\mathcal{S}}$ is the desired ideal strategy.

Ideal strategy $\hat{\mathcal{S}}$ is close to \mathcal{S}

It remains to be shown that the transformation's three steps incur a small error: $\hat{\mathcal{S}}$ is close to \mathcal{S} . A major theme in the analysis is to leverage the known tensor-product structure between \mathcal{H}_A and \mathcal{H}_B to extract a tensor-product structure within each of \mathcal{H}_A and \mathcal{H}_B .

Step 1: $\mathcal{S} \approx \tilde{\mathcal{S}}$. Although elementary, explaining this step is useful for establishing some notation. Let ρ_1 be the devices' initial shared state, possibly entangled with the environment. Let \mathcal{E}_j^A and \mathcal{E}_j^B be the super-operators that implement Alice and Bob's respective strategies for game j , let $\mathcal{E}_j^{AB} = \mathcal{E}_j^A \otimes \mathcal{E}_j^B$ and let $\mathcal{E}_{j,k}^{AB} = \mathcal{E}_k^{AB} \cdots \mathcal{E}_j^{AB}$ for $j \leq k$; thus, the state ρ_j of equation (1) equals $\mathcal{E}_{1,j-1}^{AB}(\rho_1)$. For $D \in \{A, B\}$, let $\tilde{\mathcal{E}}_j^D$ be the super-operator in which the actual measurement operators in \mathcal{E}_j^D are replaced with the ideal operators that follow from the CHSH rigidity theorem. Strategy $\tilde{\mathcal{S}}$ is given by $\rho_1, \{\tilde{\mathcal{E}}_j^A\}$ and $\{\tilde{\mathcal{E}}_j^B\}$. If $\Pr[\text{game } j \text{ is } \varepsilon\text{-structured}] \geq 1 - \delta$, then $\|\mathcal{E}_j^{AB}(\rho_j) - \tilde{\mathcal{E}}_j^{AB}(\rho_j)\|_{\text{tr}} \leq$

$2\delta + O(\sqrt{\varepsilon})$. (This expression combines bounds on the probability of the bad event and the $O(\sqrt{\varepsilon})$ error from the good event.) To achieve

the goal, namely showing that $\mathcal{E}_{1,n}^{AB}(\rho_1) \approx \tilde{\mathcal{E}}_{1,n}^{AB}(\rho_1)$ in trace distance, work backwards from game n to game 1 fixing each game's measurement operators one at a time, accumulating an error of $n(2\delta + O(\sqrt{\varepsilon}))$. Step 2: $\tilde{\mathcal{S}} \approx \hat{\mathcal{S}}$. The key to showing that $\tilde{\mathcal{S}}$ is close to $\hat{\mathcal{S}}$ is the fact that operations on one half of an EPR state can equivalently be performed on the other half, because for any 2×2 matrix M , $(M \otimes I)(|00\rangle + |11\rangle) = (I \otimes M^T)(|00\rangle + |11\rangle)$. This means that the outcome of an ε -structured CHSH game would be nearly unchanged if Bob were hypothetically to perform Alice's measurement before his own. Once Alice's measurement operators for games $j+1$ to n are moved over to Bob's side, they cannot affect the qubit $|\alpha_j\rangle$ from game j on her side. Therefore, undoing the original change of basis restores the ancilla qubits nearly to their initial state $|0^n\rangle$, and $\tilde{\mathcal{S}} \approx \hat{\mathcal{S}}$.

In more detail, define a unitary super-operator \mathcal{V}_j that rotates the j th ancilla qubit to $|\alpha_j\rangle$, depending on Alice's transcript h_j^A . Define a unitary super-operator \mathcal{T}_j to apply \mathcal{V}_j and swap the j th ancilla qubit with the qubit Alice uses in game j (depending on h_{j-1}^A). Alice's multi-qubit ideal strategy is given by

$$\tilde{\mathcal{E}}_j^A = \mathcal{T}_{1,j-1}^{-1} \left(1_{C^n} \otimes \tilde{\mathcal{E}}_j^A \right) \mathcal{T}_{1,j-1} \quad (2)$$

We aim to show that the strategy given by $\rho_1, \{\tilde{\mathcal{E}}_j^A\}$ and $\{\tilde{\mathcal{E}}_j^B\}$ is close to $\hat{\mathcal{S}}$ up to the fixed isometry that adds $|0^n\rangle\langle 0^n|$ to the state, that is, that $|0^n\rangle\langle 0^n| \otimes \tilde{\mathcal{E}}_{1,n}^{AB}(\rho_1) \approx \tilde{\mathcal{E}}_{1,n}^A(|0^n\rangle\langle 0^n| \otimes \tilde{\mathcal{E}}_{1,n}^B(\rho_1))$. Define a super-operator $\tilde{\mathcal{F}}_j^{AB}$, in which Alice's measurements are made on Bob's Hilbert space \mathcal{H}_B , on the qubit determined by Bob's transcript h_{j-1}^B . Because most games are ε -structured, it follows from the CHSH rigidity theorem that $\tilde{\mathcal{F}}_{j+1,k}^{AB}(\tilde{\rho}_{j+1}) \approx \tilde{\mathcal{E}}_{j+1,k}^{AB}(\tilde{\rho}_{j+1}) = \tilde{\rho}_{k+1}$ for $j \leq k$. Because $\tilde{\mathcal{F}}_{j+1,k}^{AB}$ acts on \mathcal{H}_B , it does not affect Alice's qubit $|\alpha_j\rangle$ from game j at all, and so this qubit must stay near $|\alpha_j\rangle$ in $\tilde{\rho}_{k+1}$ as well; that is, the trace of the reduced density matrix against the projection $|\alpha_j\rangle\langle \alpha_j|$ stays close to one. Because this holds for every j , $\mathcal{T}_{1,n}^{-1}$ indeed returns the ancillas almost to their initial state $|0^n\rangle$. The $\{\tilde{\mathcal{E}}_j^B\}$ are symmetrically adjusted to $\{\tilde{\mathcal{E}}_j^A\}$.

Step 3: $\tilde{\mathcal{S}} \approx \hat{\mathcal{S}}$. In $\tilde{\mathcal{S}}$, Alice and Bob play according to a strategy in which every game uses a qubit in tensor product with the previous games' qubits. However, the qubit's location can depend on previous games' outcomes. We wish to argue that Alice and Bob must play using a single set of n qubits, fixed in advance independent of the transcript.

Intuitively, if the location of Alice's j th qubit depended on h_{j-1}^A , then because the devices cannot communicate with each other, Bob could not know which of his qubits to measure. However, Alice and Bob's transcripts are significantly correlated, and we must show that they cannot use these correlations to coordinate dynamically the locations of their qubits.

For a toy example that illustrates the issue, consider two devices who play the first $n-1$ games honestly and which at the beginning of the last game share two EPR states, $|\phi\rangle^{\otimes 2}$. Say that for certain functions f and g , Alice uses EPR state $f(h_{n-1}^A) \in \{0, 1\}$ in game n , and Bob uses EPR state $g(h_{n-1}^B) \in \{0, 1\}$. For game n to be structured, they need $f(h_{n-1}^A) = g(h_{n-1}^B)$ so that they measure the same EPR state. Now Alice and Bob's local transcripts are each uniformly random, separately, but corresponding bits have a constant correlation. To coordinate non-trivially, the best they can do is to set f and g both to the majority function³⁷. Even then, though, $\Pr[f(h_{n-1}^A) \neq g(h_{n-1}^B)]$ would be too large. By considering the influences of each input bit on f and g , we can argue that the functions must be nearly constant. Thus, one of the two EPR states is used almost always.

This example gives an essentially classical cheating strategy. The actual devices may be significantly more sophisticated. In particular, small amounts of cheating in earlier games might enable an avalanche of more and more blatant cheating in later games, drastically changing the underlying quantum state. If, for example, Alice knowingly

manages to swap her halves of the two last EPR states along some transcripts h_{n-1}^A , then she can use completely different strategies for the last game without having to coordinate with Bob. We control such errors, as in the arguments sketched above, by replacing Alice's super-operator with one acting on Bob's side; locality then isolates the effects of errors. More formal arguments are deferred to the Supplementary Information.

Scheme for verified quantum dynamics

Our scheme for verified quantum dynamics is based on the idea of computation by teleportation, which reduces computation to preparing certain resource states and applying Bell measurements³⁰ (Fig. 2f). Say that Eve wants to simulate a quantum circuit C , over the gate set $\{H, G, \text{CNOT}\}$, where H is the Hadamard gate, $G = \exp(-i\pi\sigma_y/8)$ and CNOT is the controlled NOT. Eve asks Bob to prepare for Alice many copies of $|0\rangle \otimes (I \otimes H)|\varphi\rangle \otimes (I \otimes G)|\varphi\rangle \otimes \text{CNOT}_{2,4}(|\varphi\rangle \otimes |\varphi\rangle)$, where $|\varphi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$. He can do so by applying one-, two- and four-qubit measurements to his halves of the shared EPR states and reporting the results to Eve. If he plays honestly, Alice's shares of the EPR states collapse into the desired resource states, up to simple corrections. Each resource state corresponds to a basic operation in C . Eve wires these up by repeatedly directing Alice to make a Bell measurement connecting the output of one operation to the input of the next operation in C . After each G gate, an H correction might be required.

Of course, Alice and Bob might not follow directions. To enforce honest play, Eve runs this protocol only a small fraction of the time, and otherwise chooses uniformly between three alternative protocols sketched in Fig. 2. Let $m = |C|^{O(1)}$ and $n = m^{O(1)}$.

Protocol 1. In the 'state tomography' protocol, Eve chooses K uniformly from $\{1, \dots, n/m\}$. She referees $K-1$ blocks of m CHSH games. Then, in the K th block of m games, Eve asks Bob to prepare the resource states, in a random order, while continuing to play CHSH games with Alice. Eve rejects if the tomography statistics are inconsistent; for each multi-qubit Pauli operator, the number of measurement outcomes reported by Alice should be close to its expected value for honest play. We prove that if Alice plays honestly and Eve accepts with high probability, then on most randomly chosen small subsets of the resource state positions, Alice's reduced state before her measurements is close to the correct tensor product of resource states.

Protocol 2. In the 'process tomography' protocol, Eve again chooses K uniformly from $\{1, \dots, n/m\}$ and referees $K-1$ blocks of m CHSH games. In the K th block of m games, Eve asks Alice to make Bell measurements on random pairs of qubits, while continuing to play CHSH games with Bob. If Alice's reported result for any pair of qubits is inconsistent with Bob's outcomes, Eve rejects. Then, if Bob plays honestly and Eve accepts with high probability, Alice must also have applied the Bell measurements honestly.

Protocol 3. In this protocol, Eve simply referees n sequential CHSH games with both devices and rejects if they do not win at least $(1 - \varepsilon)\omega^*n$ games.

From Bob's perspective, the process tomography and computation protocols are indistinguishable, as are the state tomography and CHSH game protocols. From Alice's perspective, the state tomography and computation protocols are indistinguishable, as are the process tomography and CHSH game protocols. The devices must behave identically in indistinguishable protocols. The multi-game rigidity theorem therefore provides the base for a chain of implications which implies that if Eve accepts with high probability, then the devices must implement C honestly.

Four main technical problems obstruct these claims. First, in the state tomography protocol, if Bob is dishonest then Alice gets an arbitrary m -qubit state, and there is no reason why it should split into a tensor product of repeated, constant-qubit states. Nonetheless, we argue using martingales that if the counts of Alice's different measurement outcomes roughly match their expectations with high probability, then for most reported measurement outcomes from Bob and for most subsystems j , Alice's conditional state reduced to her j th subsystem is close to what it should be.

Second, saturating Tsirelson's inequality for the CHSH game implies only that Alice is honestly making Pauli σ_x and σ_z measurements on her half of an EPR state. Tomography also requires σ_y measurements. To sidestep this issue, we generalize a theory of ref. 38 and prove that there is a large class of states, including the necessary resource states, that are all robustly determined by only σ_x and σ_z measurements.

A third and bigger problem, though, is that we want to characterize the operations that each device applies to the shared EPR states, and not just the states that these operations create on the other device's

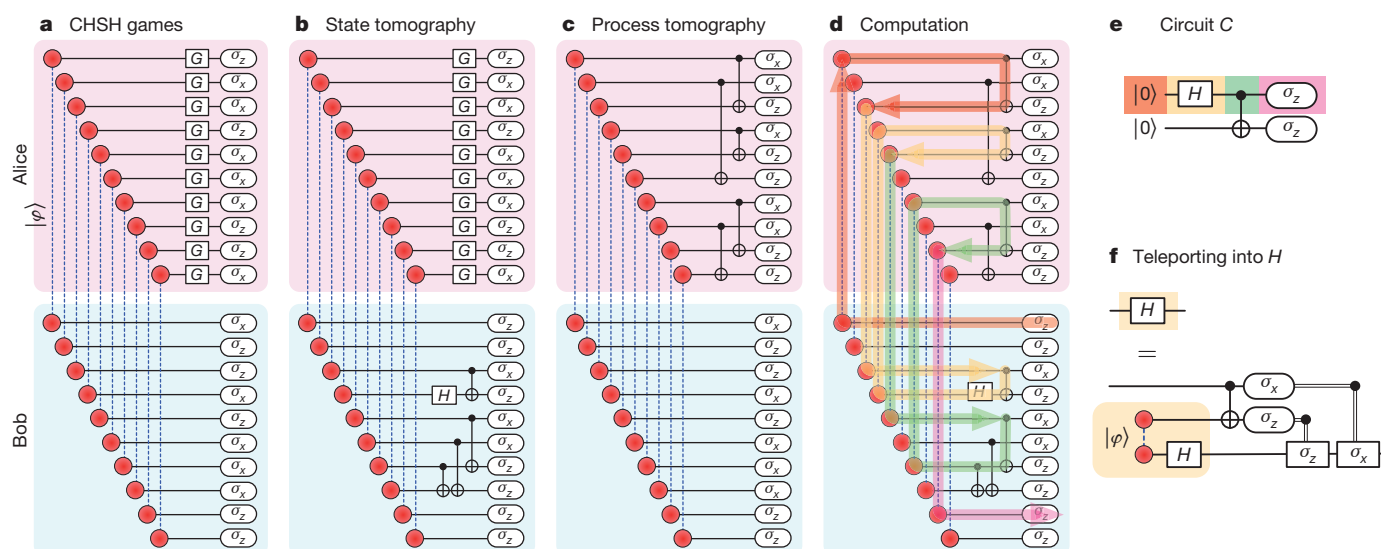


Figure 2 | Sub-protocols for verified quantum dynamics. To delegate a quantum computation, Eve runs a random one of four sub-protocols with Alice (top row, a–d) and Bob (bottom row, a–d). **a**, Playing many CHSH games ensures that the devices play honestly, measuring in each game an EPR state $|\varphi\rangle$ on two qubits (red dots). **b, c**, This lets Eve apply state (b) or process (c) tomography to characterize more complicated multi-qubit operations.

d, e, By adaptively combining these operations (d), Eve directs a quantum circuit C (e). The operations along the zig-zagging logical path of the first qubit of C are in d highlighted using the same colours as in e. **f**, Each gate of C is implemented through teleportation; in this simpler example, H is applied by a Bell measurement on half of the resource state $(I \otimes H)|\varphi\rangle$.

side. The distinction is the same as that between process and state tomography. Essentially, the problem is that the correct states could be generated by incorrect processes. Moreover, as for sequential CHSH games, Bob's strategy in early tomography rounds might be sufficiently dishonest as to allow him in later rounds to apply completely dishonest operators. A key observation to avoid this problem is that it is enough to certify the states prepared by one device and the processes applied by the other. Then, because a broad class of states can be certified, for applications it suffices to certify a much smaller set of operations. We restrict consideration to Pauli stabilizer measurements³⁹. For Pauli operators in the stabilizer of a state, the measurement outcome is deterministic. Therefore, if Alice reports the wrong stabilizer syndrome in even a single round, Eve can reject. Our process certification analysis is similar to the arguments used in step 2 of the proof of the multi-game rigidity theorem. We argue that Alice's earlier measurements cannot usually overly disturb the qubits intended for use in later measurements, by moving Alice's measurement super-operators over onto Bob's halves of the EPR states.

Finally, the verifier's questions in the state and process tomography protocols are non-adaptive, whereas in computation by teleportation the questions must be chosen adaptively on the basis of previous responses. This is an attack vector in some related protocols³⁶. However, we argue that the devices can learn nothing from the adaptive questions. This follows because computation by teleportation can be implemented exactly equivalently either by choosing Bob's state preparation questions non-adaptively and Alice's process questions adaptively, or vice versa.

The proof that QMIP = MIP* follows along similar lines. Begin with a k -prover protocol with a quantum verifier. We may assume that there are two rounds of quantum messages from the provers, one before and one after the verifier broadcasts a random bit⁴⁰. To convert to a protocol with a classical verifier, Eve, add two new provers, Alice and Bob. Eve teleports the original k provers' messages to Alice, and directs Alice and Bob together to apply the quantum verifier's acceptance predicate.

Discussion

By characterizing the device strategies that can win many successive CHSH games, we have shown how a fully classical party can direct the actions of two untrusted quantum devices. The simplest case is DIQKD, free of the independence assumptions needed in previous analyses. Following the pattern established in refs 9, 10, the QKD devices begin with shared entanglement and the two experimentalists act together as 'Eve'. They gather statistics as in the verified computation protocol to certify the devices' shared state and measurement operators, and extract secret key material from a random block of games. Two major challenges are to improve the efficiency of the scheme, to get a constant key rate instead of inverse-polynomial in n , and to tolerate a constant noise rate. More generally, the CHSH multi-game rigidity theorem may be viewed as a quantum analogue of classical multilinearity tests, which are central to the theory of probabilistically checkable proofs; by simple local tests, it guarantees the existence of a special type of large quantum state.

Received 6 September 2012; accepted 19 February 2013.

- Aharonov, D., Ben-Or, M. & Eban, E. in *Proc. Innovations in Computer Science (ICS)* (ed. Yao, A.) 453–469 (Tsinghua Univ. Press, 2010).
- Broadbent, A., Fitzsimons, J. F. & Kashefi, E. in *Proc. IEEE Foundations of Computer Science (FOCS)* 517–526 (IEEE Computer Society, 2009).
- Bennett, C. H. & Brassard, G. in *Proc. IEEE Int. Conf. on Computers, Systems and Signal Processing* 175–179 (IEEE Computer Society, 1984).
- Lo, H.-K. & Chau, H. F. Unconditional security of quantum key distribution over arbitrarily long distances. *Science* **283**, 2050–2056 (1999).
- Shor, P. W. & Preskill, J. Simple proof of security of the BB84 quantum key distribution protocol. *Phys. Rev. Lett.* **85**, 441–444 (2000).
- Zhao, Y., Fung, C.-H. F., Qi, B., Chen, C. & Lo, H.-K. Quantum hacking: experimental demonstration of time-shift attack against practical quantum key distribution systems. *Phys. Rev. A* **78**, 042333 (2008).
- Lydersen, L. *et al.* Hacking commercial quantum cryptography systems by tailored bright illumination. *Nature Photon.* **4**, 686–689 (2010).
- Gerhardt, I. *et al.* Full-field implementation of a perfect eavesdropper on a quantum cryptography system. *Nature Commun.* **2**, 349 (2011).
- Mayers, D. & Yao, A. in *Proc. IEEE Foundations of Computer Science (FOCS)* 503–509 (IEEE Computer Society, 1998).
- Ekert, A. K. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.* **67**, 661–663 (1991).
- Barrett, J., Hardy, L. & Kent, A. No signalling and quantum key distribution. *Phys. Rev. Lett.* **95**, 010503 (2005).
- Masanes, L., Renner, R., Christandl, M., Winter, A. & Barrett, J. Unconditional security of key distribution from causality constraints. Preprint at <http://arXiv.org/abs/quant-ph/0606049> (2006).
- Acín, A., Massar, S. & Pironio, S. Efficient quantum key distribution secure against no-signalling eavesdroppers. *N. J. Phys.* **8**, 126 (2006).
- Masanes, L. Universally composable privacy amplification from causality constraints. *Phys. Rev. Lett.* **102**, 140501 (2009).
- Hänggi, E., Renner, R. & Wolf, S. in *Proc. EUROCRYPT* (ed. Gilbert, H.) 216–234 (LNCS 6110, Springer, 2010).
- Acín, A. *et al.* Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.* **98**, 230501 (2007).
- Pironio, S. *et al.* Device-independent quantum key distribution secure against collective attacks. *N. J. Phys.* **11**, 045021 (2009).
- McKague, M. Device independent quantum key distribution secure against coherent attacks with memoryless measurement devices. *N. J. Phys.* **11**, 103037 (2009).
- Hänggi, E. & Renner, R. Device-independent quantum key distribution with commuting measurements. Preprint at <http://arXiv.org/abs/1009.1833> (2010).
- Masanes, L., Pironio, S. & Acín, A. Secure device-independent quantum key distribution with causally independent measurement devices. *Nature Commun.* **2**, 238 (2011).
- Holevo, A. S. Bounds for the quantity of information transmitted by a quantum communication channel. *Probl. Inf. Transm.* **9**, 177–183 (1973).
- Bell, J. S. On the Einstein-Podolsky-Rosen paradox. *Physics* **1**, 195–200 (1964).
- Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**, 880–884 (1969).
- Cirel'son, B. S. Quantum generalizations of Bell's inequality. *Lett. Math. Phys.* **4**, 93–100 (1980).
- Braunstein, S. L., Mann, A. & Revzen, M. Maximal violation of Bell inequalities for mixed states. *Phys. Rev. Lett.* **68**, 3259–3261 (1992).
- Popescu, S. & Rohrlich, D. Which states violate Bell's inequality maximally? *Phys. Lett. A* **169**, 411–414 (1992).
- Magniez, F., Mayers, D., Mosca, M. & Olivier, H. in *Proc. Int. Coll. on Automata, Languages and Programming (ICALP)* (eds Bugliesi, M. *et al.*) 72–83 (LNCS 4051, Springer, 2006).
- McKague, M., Yang, T. H. & Scarani, V. Robust self-testing of the singlet. *J. Phys. A* **45**, 455304 (2012).
- Miller, C. & Shi, Y. Robust self-testing quantum states and binary nonlocal XOR games. Preprint at <http://arXiv.org/abs/1207.1819> (2012).
- Gottesman, D. & Chuang, I. L. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature* **402**, 390–393 (1999).
- Watrous, J. PSPACE has constant-round quantum interactive proof systems. *Theor. Comput. Sci.* **292**, 575–588 (2003).
- Fitzsimons, J. F. & Kashefi, E. Unconditionally verifiable blind computation. Preprint at <http://arXiv.org/abs/1203.5217> (2012).
- Barz, S. *et al.* Demonstration of blind quantum computing. *Science* **335**, 303–308 (2012).
- Broadbent, A., Fitzsimons, J. F. & Kashefi, E. QMIP = MIP*. Preprint at <http://arXiv.org/abs/1004.1130> (2010).
- Jordan, C. Essai sur la géométrie à n dimensions. *Bull. Soc. Math. Fr.* **3**, 103–174 (1875).
- Reichardt, B. W., Unger, F. & Vazirani, U. A classical leash for a quantum system: command of quantum systems via rigidity of CHSH games. Preprint at <http://arXiv.org/abs/1209.0448> (2012).
- Mossel, E., O'Donnell, R. & Oleszkiewicz, K. Noise stability of functions with low influences: invariance and optimality. *Ann. Math.* **171**, 295–341 (2010).
- McKague, M. *Quantum Information Processing with Adversarial Devices*. PhD thesis, Univ. Waterloo (2010).
- Gottesman, D. *Stabilizer Codes and Quantum Error Correction*. PhD thesis, California Inst. Technol. (1997).
- Kempe, J., Kobayashi, H., Matsumoto, K. & Vidick, T. Using entanglement in quantum multi-prover interactive proofs. *Comput. Complex.* **18**, 273–307 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank E. Bering, A. Broadbent, A. Chailloux, M. Christandl, R. Colbeck, T. Ito, R. König, M. McKague, V. Madhavan, R. Renner, S. Sondhi and T. Vidick for discussions. Part of this work was conducted while F.U. was at the University of California Berkeley and B.W.R. was at the Institute for Quantum Computing, University of Waterloo. B.W.R. acknowledges support from NSERC, ARO-DTO and Mitacs. U.V. acknowledges support from US NSF grant CCF-0905626 and Templeton grant 21674.

Author Contributions All authors made significant contributions to the research presented in this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.W.R. (ben.reichardt@usc.edu).

Dynamic regulatory network controlling T_H17 cell differentiation

Nir Yosef^{1,2*}, Alex K. Shalek^{3*}, Jellert T. Gaublomme^{3*}, Hulin Jin², Youjin Lee², Amit Awasthi^{2†}, Chuan Wu², Katarzyna Karwacz², Sheng Xiao², Marsela Jorgolli³, David Gennert¹, Rahul Satija¹, Arvind Shakya⁴, Diana Y. Lu¹, John J. Trombetta¹, Meenu R. Pillai⁵, Peter J. Ratcliffe⁶, Mathew L. Coleman⁶, Mark Bix⁵, Dean Tantin⁴, Hongkun Park^{1,3}, Vijay K. Kuchroo^{1,2} & Aviv Regev^{1,7}

Despite their importance, the molecular circuits that control the differentiation of naive T cells remain largely unknown. Recent studies that reconstructed regulatory networks in mammalian cells have focused on short-term responses and relied on perturbation-based approaches that cannot be readily applied to primary T cells. Here we combine transcriptional profiling at high temporal resolution, novel computational algorithms, and innovative nanowire-based perturbation tools to systematically derive and experimentally validate a model of the dynamic regulatory network that controls the differentiation of mouse T_H17 cells, a proinflammatory T-cell subset that has been implicated in the pathogenesis of multiple autoimmune diseases. The T_H17 transcriptional network consists of two self-reinforcing, but mutually antagonistic, modules, with 12 novel regulators, the coupled action of which may be essential for maintaining the balance between T_H17 and other CD4⁺ T-cell subsets. Our study identifies and validates 39 regulatory factors, embeds them within a comprehensive temporal network and reveals its organizational principles; it also highlights novel drug targets for controlling T_H17 cell differentiation.

Effective coordination of the immune system requires careful balancing of distinct pro-inflammatory and regulatory CD4⁺ helper T-cell populations. Among those, pro-inflammatory IL-17 producing T_H17 cells have a key role in the defence against extracellular pathogens and have also been implicated in the induction of several autoimmune diseases¹. T_H17 differentiation from naive T cells can be triggered *in vitro* by the cytokines TGF- β 1 and IL-6. Whereas TGF- β 1 alone induces Foxp3⁺ regulatory T cells (T_{reg} cells)², the presence of IL-6 inhibits T_{reg} development and induces T_H17 differentiation¹.

Much remains unknown about the regulatory network that controls the differentiation of T_H17 cells^{3,4}. Developmentally, as TGF- β is required for both T_H17 and induced T_{reg} differentiation, it is not fully understood how balance is achieved between them or how IL-6 produces a bias towards T_H17 differentiation¹. Functionally, it is unclear how the pro-inflammatory status of T_H17 cells is held in check by the immunosuppressive cytokine IL-10 (refs 3, 4). Finally, many of the key regulators and interactions that drive the development of T_H17 cells remain unknown⁵.

Recent studies have demonstrated the power of coupling systematic profiling with perturbation for deciphering mammalian regulatory circuits^{6–9}. Most of these studies have relied upon computational circuit-reconstruction algorithms that assume one ‘fixed’ network. T_H17 differentiation, however, spans several days, during which the components and wiring of the regulatory network probably change. Furthermore, naive T cells and T_H17 cells cannot be transfected effectively *in vitro* by traditional methods without changing their phenotype or function, thus limiting the effectiveness of perturbation strategies for inhibiting gene expression.

Here we address these limitations by combining transcriptional profiling, novel computational methods and nanowire-based short

interfering RNA (siRNA) delivery¹⁰ (Fig. 1a) to construct and validate the transcriptional network of T_H17 differentiation. The reconstructed model is organized into two coupled, antagonistic and densely intra-connected modules, one promoting and the other suppressing the T_H17 program. The model highlights 12 novel regulators, the function of which we further characterized by their effects on global gene expression, DNA binding profiles, or T_H17 differentiation in knockout mice.

A transcriptional time course of T_H17 differentiation

We induced the differentiation of naive CD4⁺ T cells into T_H17 cells using TGF- β 1 and IL-6, and measured transcriptional profiles using microarrays at 18 time points along a 72-h time course (Fig. 1, Supplementary Fig. 1a–c and Methods). As controls, we measured mRNA profiles for cells that were activated without the addition of differentiating cytokines (T_H0). We identified 1,291 genes that were differentially expressed specifically during T_H17 differentiation (Methods and Supplementary Table 1) and partitioned them into 20 co-expression clusters (*k*-means clustering; Methods, Fig. 1b and Supplementary Fig. 2) with distinct temporal profiles. We used these clusters to characterize the response and reconstruct a regulatory network model, as described below (Fig. 2a).

Three main waves of transcription and differentiation

There are three transcriptional phases as the cells transition from a naive-like state ($t = 0.5$ h) to T_H17 ($t = 72$ h; Fig. 1c and Supplementary Fig. 1c): early (up to 4 h), intermediate (4–20 h), and late (20–72 h). Each corresponds, respectively, to a differentiation phase⁵: (1) induction; (2) onset of phenotype and amplification; and (3) stabilization and IL-23 signalling. The early phase is characterized

¹Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ²Center for Neurologic Diseases, Brigham & Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ³Department of Chemistry and Chemical Biology and Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA. ⁴Department of Pathology, University of Utah School of Medicine, Salt Lake City, Utah 84132, USA. ⁵Department of Immunology, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ⁶University of Oxford, Headington Campus, Oxford OX3 7BN, UK. ⁷Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02140, USA. [†]Present address: Translational Health Science & Technology Institute, Faridabad, Haryana 122016, India.

*These authors contributed equally to this work.

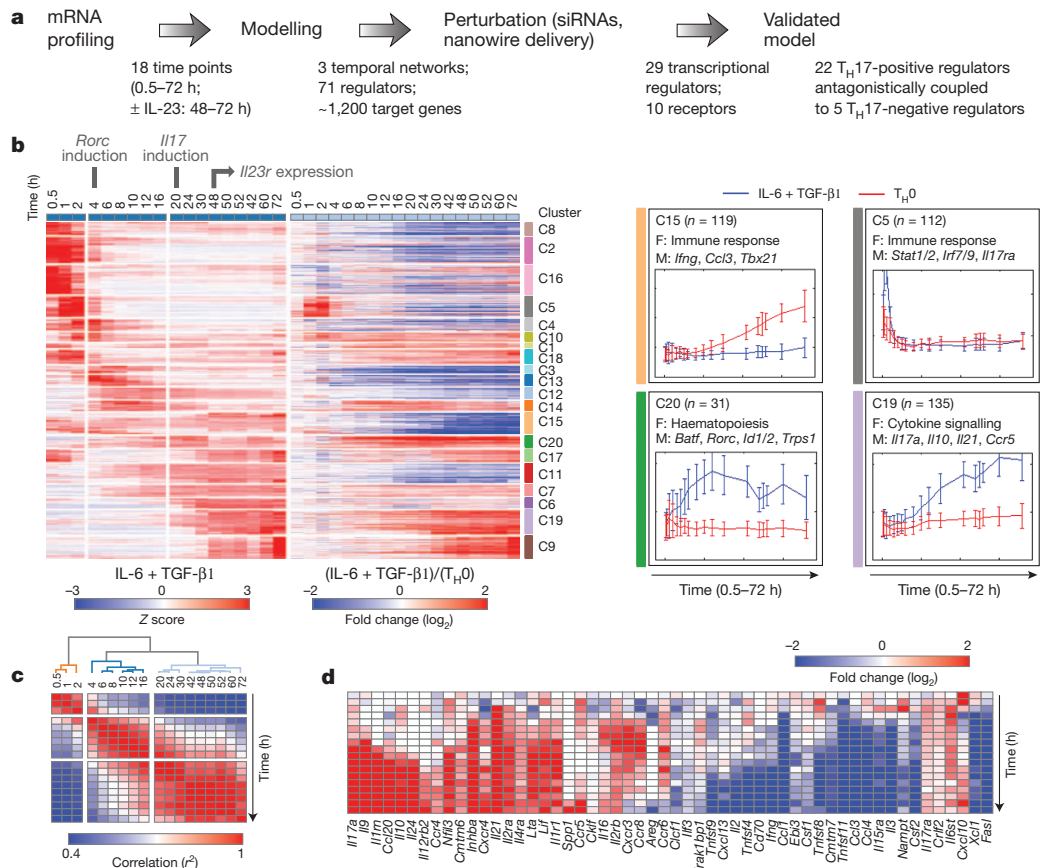


Figure 1 | Genome-wide temporal expression profiles of T_H17 differentiation. **a**, Overview of approach. **b**, Gene expression profiles during T_H17 differentiation. Shown are the differential expression levels for genes (rows) at 18 time points (columns) in T_H17 polarizing conditions (TGF- β 1 and IL-6; left panel, Z-normalized per row) or T_H17 polarizing conditions relative to control activated T_H0 cells (right panel, $\log_2(\text{ratio})$). The genes are partitioned into 20 clusters (C1–C20, colour bars, right). Right: mean expression (y axis)

by transient induction (for example, cluster C5, Fig. 1b) of immune response pathways (for example, IL-6 and TGF- β signalling; Supplementary Table 2). Some early induced genes display sustained expression (for example, cluster C10, Fig. 1b); these are enriched for transcription factors, including the key T_H17 factors *Stat3*, *Irf4* and *Batf*, and the cytokine and cytokine receptors *Il21*, *Lif* and *Il2ra* (Supplementary Table 1). The transition to the intermediate phase ($t = 4$ h) is marked by induction of the *Rorc* gene (encoding the master transcription factor ROR- γ t; Supplementary Fig. 1d) and another 12 transcription factors (cluster C20, Fig. 1b), both known (for example, *Ahr*) and novel (for example, *Trps1*) in T_H17 differentiation. During the transition to the late phase ($t = 20$ h), mRNAs of T_H17 signature cytokines are induced (for example, *Il17a*, *Il9*; cluster C19) whereas mRNAs of cytokines that signal other T-cell lineages are repressed (for example, *Ifng* and *Il4*). Regulatory cytokines from the IL-10 family are also induced (*Il10*, *Il24*), possibly as a self-limiting mechanism related to the emergence of ‘pathogenic’ or ‘non-pathogenic’ T_H17 cells¹¹. Around 48 h, the cells induce *Il23r* (data not shown), which has an important role in the late phase (Supplementary Fig. 3 and Supplementary Table 1).

Inference of dynamic regulatory interactions

We proposed the hypothesis that each of the clusters (Fig. 1b and Supplementary Table 2) encompasses genes that share regulators active in the relevant time points. To predict these regulators, we assembled a general network of regulator–target associations from published genomics profiles^{12–19} (Fig. 2a and Methods). We then

connected a regulator to a gene from its set of putative targets only if there was also a significant overlap between the regulator’s putative targets and that gene’s cluster (Methods). Because different regulators act at different times, the connection between a regulator and its target may be active only within a certain time window. To determine this window, we labelled each edge with a time stamp denoting when both the target gene is regulated (based on its expression profile) and the regulator node is expressed at sufficient levels (based on its mRNA levels and inferred protein levels²⁰; Methods). In this way, we derived a network ‘snapshot’ for each of the 18 time points (Fig. 2b–d). Overall, 10,112 interactions between 71 regulators and 1,283 genes were inferred in at least one network.

connected a regulator to a gene from its set of putative targets only if there was also a significant overlap between the regulator’s putative targets and that gene’s cluster (Methods). Because different regulators act at different times, the connection between a regulator and its target may be active only within a certain time window. To determine this window, we labelled each edge with a time stamp denoting when both the target gene is regulated (based on its expression profile) and the regulator node is expressed at sufficient levels (based on its mRNA levels and inferred protein levels²⁰; Methods). In this way, we derived a network ‘snapshot’ for each of the 18 time points (Fig. 2b–d). Overall, 10,112 interactions between 71 regulators and 1,283 genes were inferred in at least one network.

Substantial regulatory re-wiring during differentiation

The active factors and interactions change from one network to the next. The vast majority of interactions are active only at some time windows (Fig. 2c), even for regulators (for example, *Batf*) that participate in all networks. On the basis of similarity in active interactions, we identified three network classes (Fig. 2c) corresponding to the three differentiation phases (Fig. 2d). We collapsed all networks in each phase into one model, resulting in three consecutive network models (Fig. 2d, Supplementary Fig. 4 and Supplementary Table 3). Among the regulators, 33 are active in all of the networks (for example, known master regulators such as *Batf*, *Irf4* and *Stat3*), whereas 18 are active primarily in one (for example, *Stat1* and *Irf1* in the early network; ROR- γ t in the late network). Indeed, whereas *Rorc* mRNA levels are induced at ~ 4 h, ROR- γ t protein levels increase

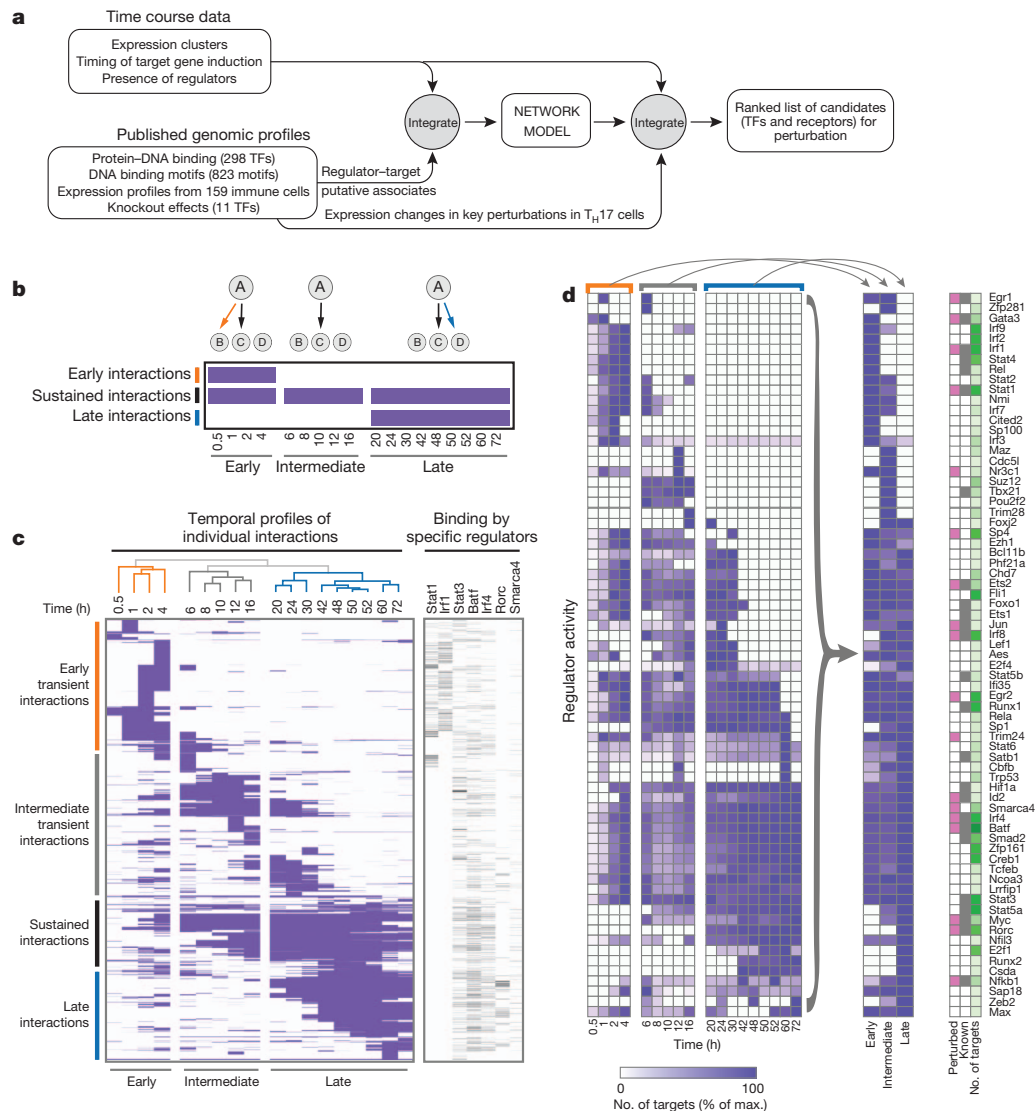


Figure 2 | A model of the dynamic regulatory network of T_H17 differentiation. **a**, Overview of computational analysis. **b**, Schematic of temporal network 'snapshots'. Shown are three consecutive cartoon networks (top) and matrix columns (bottom), with three possible interactions from regulator (A) to targets (B, C and D), shown as edges (top) and matrix rows (A→B, top row; A→C, middle row; A→D, bottom row). **c**, Eighteen network 'snapshots'. Left: each row corresponds to a transcription factor (TF)-target interaction that occurs in at least one network; columns correspond to the network at each time point. A purple entry indicates that an interaction is active in that network. The networks are clustered by similarity of active interactions (dendrogram,

at approximately 20 h and further rise over time, consistent with our model (Supplementary Fig. 5).

Ranking novel regulators for systematic perturbation

In addition to known T_H17 regulators, our network includes dozens of novel factors as predicted regulators (Fig. 2d), induced target genes, or both (Supplementary Fig. 4 and Supplementary Table 3). It also contains receptor genes as induced targets, both previously known in T_H17 cells (for example, *Il1r1*, *Il17ra*) and novel (for example, *Fas*, *Itga3*).

We ranked candidate regulators for perturbation (Figs 2a and 3a; see Methods), guided by features that reflect a regulatory role (Fig. 3a, 'Network information') and a role as a target (Fig. 3a, 'Gene expression information'). We computationally ordered the genes to emphasize certain features (for example, a predicted regulator of key T_H17 genes) over others (for example, differential expression in our time course data). We used a similar scheme to rank receptor

top), forming three temporally consecutive clusters (early, intermediate, late; bottom). Right: heat map denoting edges for selected regulators. **d**, Dynamic regulator activity. Shown is, for each regulator (rows), the number of target genes (normalized by its maximum number of targets) in each of the 18 networks (columns, left), and in each of the three canonical networks (middle) obtained by collapsing (arrows). Right: regulators chosen for perturbation (pink), known T_H17 regulators (grey), and the maximal number of target genes across the three canonical networks (green, ranging from 0 to 250 targets).

proteins (Supplementary Table 4 and Methods). Supporting their quality, our top-ranked factors are enriched ($P < 10^{-5}$) for manually curated T_H17 regulators (Supplementary Fig. 6), and correlate well (Spearman $r > 0.86$) with a ranking learned by a supervised method (Methods). We chose 65 genes for perturbation: 52 regulators and 13 receptors (Supplementary Table 4). These included most of the top 44 regulators and top 9 receptors (excluding a few well known T_H17 genes and/or those for which knockout data already existed), as well as additional representative lower ranking factors.

Nanowire-based perturbation of primary T cells

In unstimulated primary mouse T cells, viral- or transfection-based siRNA delivery has been nearly impossible because it either alters differentiation or cell viability^{21,22}. We therefore used a new delivery technology based on silicon nanowires^{10,23}, which we optimized to deliver siRNA effectively (>95%) into naive T cells without activating them (Fig. 3b, c)²³.

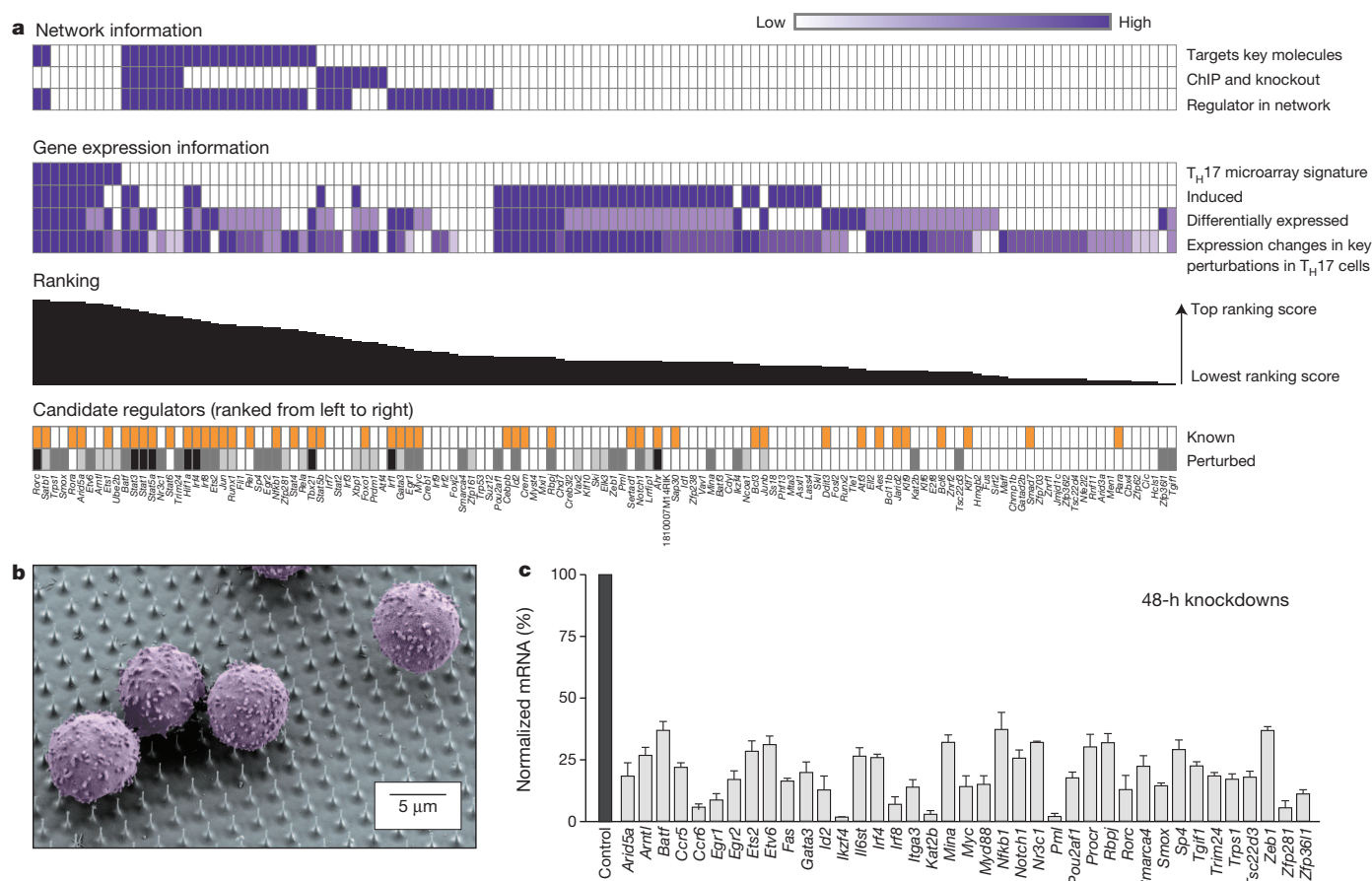


Figure 3 | Knockdown screen in T_H17 differentiation using silicon nanowires. **a**, Unbiased ranking of perturbation candidates. Shown are the genes ordered from left to right based on their ranking for perturbation (columns, top ranking is left-most). Two top matrices: criteria for ranking by ‘Network information’ (topmost) and ‘Gene expression information’. Purple entry: gene has the feature (intensity proportional to feature strength; top five features are binary). Bar chart indicates ranking score. ‘Perturbed’ row: dark grey, genes successfully perturbed by knockdown followed by high-quality mRNA quantification; light grey, genes that we attempted to knockdown but could not achieve or maintain sufficient knockdown or did not obtain enough

We attempted to perturb 60 genes with nanowire-mediated siRNA delivery and achieved efficient knockdown ($<60\%$ transcript remaining at 48 h post-activation) for 34 genes (Fig. 3c and Supplementary Fig. 7). We obtained knockout mice for seven other genes, two of which (*Irf8* and *Il17ra*) were also in the knockdown set (Supplementary Table 4). Altogether, we successfully perturbed 39 of the 65 selected genes—29 regulators and 10 receptors—including 21 genes not previously associated with T_H17 differentiation.

Nanowire-based screen validates 39 network regulators

We measured the effects of perturbations at 48 h post-activation on the expression of 275 signature genes using the Nanostring nCounter system (Supplementary Tables 5 and 6; *Il17ra* and *Il21r* knockouts were also measured at 60 h). The signature genes were computationally chosen to cover as many aspects of the differentiation process as possible (Methods): they include most differentially expressed cytokines, transcription factors, and cell-surface molecules, as well as representatives from each cluster (Fig. 1b) or enriched function (Supplementary Table 2), and predicted targets in each network (Supplementary Table 3). For validation, we profiled a signature of 86 genes using the Fluidigm BioMark system, obtaining highly reproducible results (Supplementary Fig. 8).

We scored the statistical significance of a perturbation's effect on a signature gene by comparing to non-targeting siRNAs and to 18

replicates; Black, genes that we perturbed by knockout or for which knockout data were already available. Known row: orange entry: a gene was previously associated with T_H17 function (this information was not used to rank the genes; Supplementary Fig. 6). **b**, Scanning electron micrograph of primary T cells (false-coloured purple) cultured on vertical silicon nanowires. **c**, Effective knockdown by siRNA delivered on nanowires. Shown is the percentage of mRNA remaining after knockdown (by qPCR, *y* axis: mean \pm standard error relative to non-targeting siRNA control, *n* = 12, black bar on left) at 48 h after activation.

control genes that were not differentially expressed (Supplementary Information and Fig. 4a, all non-grey entries are significant). Supporting the original network model (Fig. 2), there is a significant overlap between the genes affected by a regulator's knockdown and its predicted targets ($P \leq 0.01$, permutation test; Supplementary Information).

To study the network's dynamics, we measured the effect of 28 of the perturbations at 10 h (shortly after the induction of *Rorc*; Supplementary Table 5) using the Fluidigm BioMark system. We found that 30% of the functional interactions are present with the same activation/repression logic at both 10 h and 48 h, whereas the rest are present only in one time point (Supplementary Fig. 9). This is consistent with the extent of rewiring in our original model (Fig. 2c).

Two coupled antagonistic circuits in the T_H17 network

Characterizing each regulator by its effect on T_H17 signature genes (for example, *Il17a*, *Il17f*; Fig. 4b, grey nodes, bottom), we found that, at 48 h, the network is organized into two antagonistic modules: a module of 22 'T_H17-positive factors' (Fig. 4b, blue nodes: 9 novel), the perturbation of which decreased the expression of T_H17 signature genes (Fig. 4b, grey nodes, bottom), and a module of 5 'T_H17-negative factors' (Fig. 4b, red nodes: 3 novel), the perturbation of which had the opposite effect. Each of the modules is tightly intra-connected through positive, self-reinforcing interactions between its members

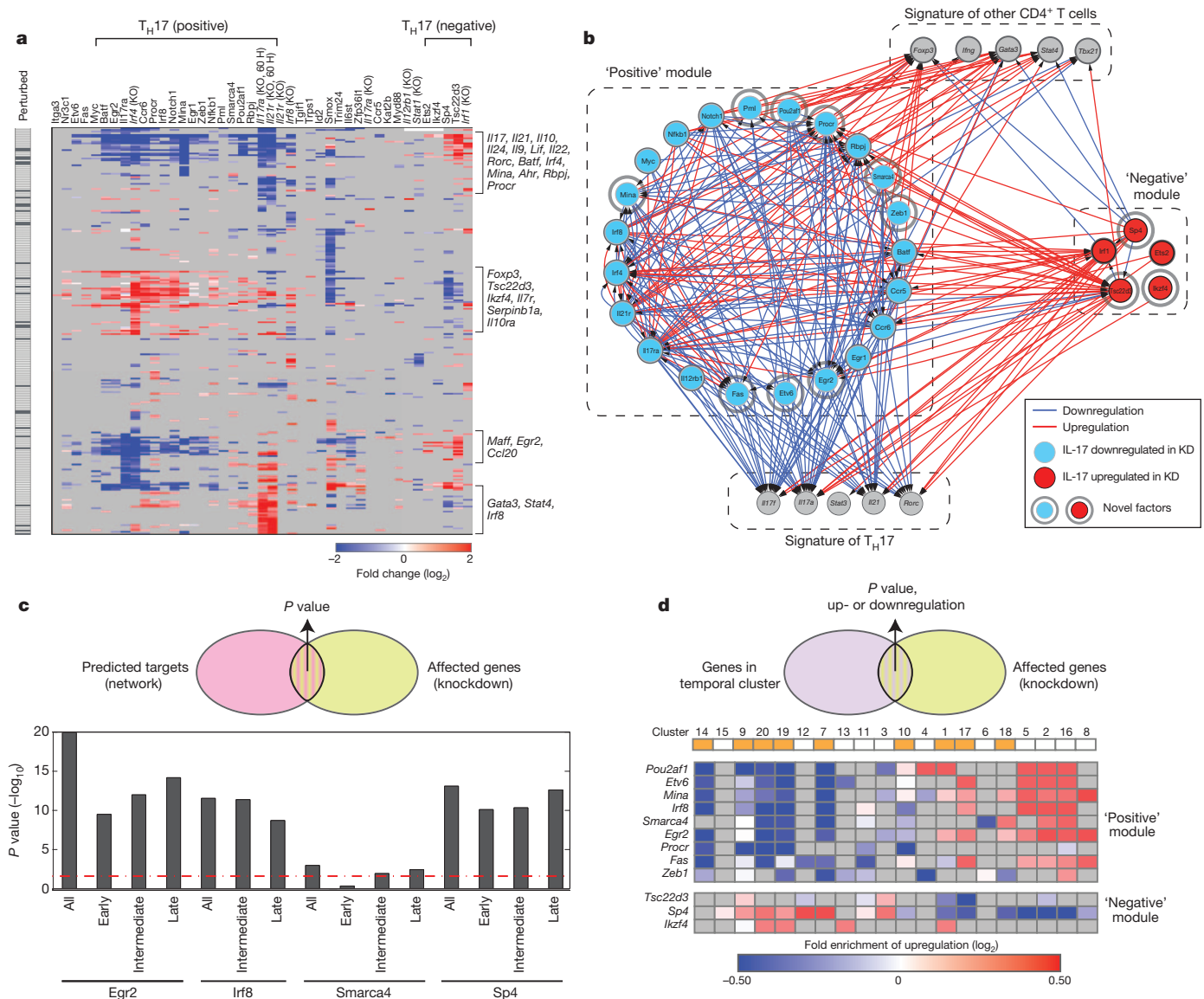


Figure 4 | Coupled and mutually antagonistic modules in the T_H17 network. **a**, Impact of perturbed genes on a 275-gene signature. Shown are changes in the expression of 275 signature genes (rows) following knockdown or knockout (KO) of 39 factors (columns) at 48 h (as well as *Il21r* and *Il17ra* knockout at 60 h). Blue, decreased expression of target following perturbation of a regulator (compared to a non-targeting control); red, increased expression; grey, not significant; all non-grey entries are significant (Supplementary Information). Perturbed (left): signature genes that are also perturbed as regulators (black entries). Key signature genes are denoted on right. **b**, Two coupled and opposing modules. Shown is the perturbation network associating the 'positive regulators' (blue nodes) of T_H17 signature genes, the 'negative regulators' (red nodes), T_H17 signature genes (grey nodes, bottom) and signature genes of other $CD4^+$ T cells (grey nodes, top). A blue edge from node A to B indicates that knockdown of A downregulates B; a red edge indicates that knockdown of A upregulates B. Light-grey halos: regulators not previously associated with T_H17 differentiation. **c**, Knockdown effects validate edges in network model. Venn diagram: we

(70% of the intra-module edges), whereas most (88%) inter-module interactions are negative. This organization, which is statistically significant (empirical P value $< 10^{-3}$; Methods, Supplementary Fig. 10), is reminiscent of that observed previously in genetic circuits in yeast^{24,25}. At 10 h, the same regulators do not yield this clear pattern ($P > 0.5$), suggesting that, at that point, the network is still malleable.

The two antagonistic modules may have a key role in maintaining the balance between T_H17 and other T-cell subsets and in self-limiting the pro-inflammatory status of T_H17 cells. Indeed, perturbing T_H17 -positive

factors also induces signature genes of other T-cell subsets, whereas perturbing T_H17 -negative factors suppresses them (for example, *Foxp3*, *Gata3* and *Stat4*; Fig. 4b, grey nodes, top).

Validation and characterization of novel factors

Next, we focused on the role of 12 of the positive or negative factors (including 11 of the 12 novel factors that have not been associated with T_H17 cells; Fig. 4b, light-grey halos). After knockdown of each factor, we used RNA-seq analysis to test whether its predicted targets

(Fig. 2) were affected (Fig. 4c, Venn diagram, top). We found highly significant overlaps ($P < 10^{-5}$) for three of the factors (*Egr2*, *Irf8* and *Sp4*) that exist in both data sets, and a borderline significant overlap for the fourth (*Smarca4*), validating the quality of the edges in our network.

Next, we assessed the designation of each of the 12 factors as 'T_H17 positive' or 'T_H17 negative' by comparing the set of genes that respond to that factor's knockdown (in RNA-seq) to each of the 20 clusters (Fig. 1b). Consistent with the original definitions, knockdown of a T_H17-positive regulator downregulated genes in otherwise induced clusters and upregulated genes in otherwise repressed or uninduced clusters (and vice versa for T_H17-negative regulators; Fig. 4d and Supplementary Fig. 11a, b). The genes affected by either positive or negative regulators also significantly overlap with those bound by key CD4⁺ transcription factors (for example, *Foxp3* (refs 26, 27), *Batf*, *Irf4* and *ROR-γt* (refs 28, 29), S. Xiao *et al.*, unpublished data).

Mina promotes the T_H17 and inhibits the Foxp3 program

Knockdown of Mina, a chromatin regulator from the Jumonji C (JmjC) family, represses the expression of signature T_H17 cytokines and transcription factors (for example, *Rorc*, *Batf*, *Irf4*) and of late-induced genes (clusters C9 and C19; $P < 10^{-5}$; Supplementary Tables 5 and 7), while increasing the expression of *Foxp3*, the master transcription factor of T_{reg} cells. Mina is strongly induced during T_H17 differentiation (cluster C7), is downregulated in *Il23r*^{-/-} T_H17 cells, and is a predicted target of *Batf*³⁰, *ROR-γt*³⁰ and *Myc* in our model (Fig. 5a). Mina was shown to suppress T_H2 bias by interacting with the transcription factor NFAT and repressing the *Il4* promoter³¹. However, in our cells, Mina knockdown did not induce T_H2 genes, indicating an alternative mode of action via positive feedback loops between Mina, *Batf* and *ROR-γt* (Fig. 5a, left). Consistent with this model, *Mina* expression is reduced in T_H17 cells from *Rorc* knockout mice, and the *Mina* promoter was found to be bound by *ROR-γt* by ChIP-seq (data not shown). Finally, the genes induced by Mina knockdown significantly overlap with those bound by *Foxp3* in T_{reg} cells^{26,27} ($P < 10^{-25}$; Supplementary Table 7) and with a cluster previously linked to *Foxp3* activity in T_{reg} cells³² (Supplementary Fig. 11c and Supplementary Table 7).

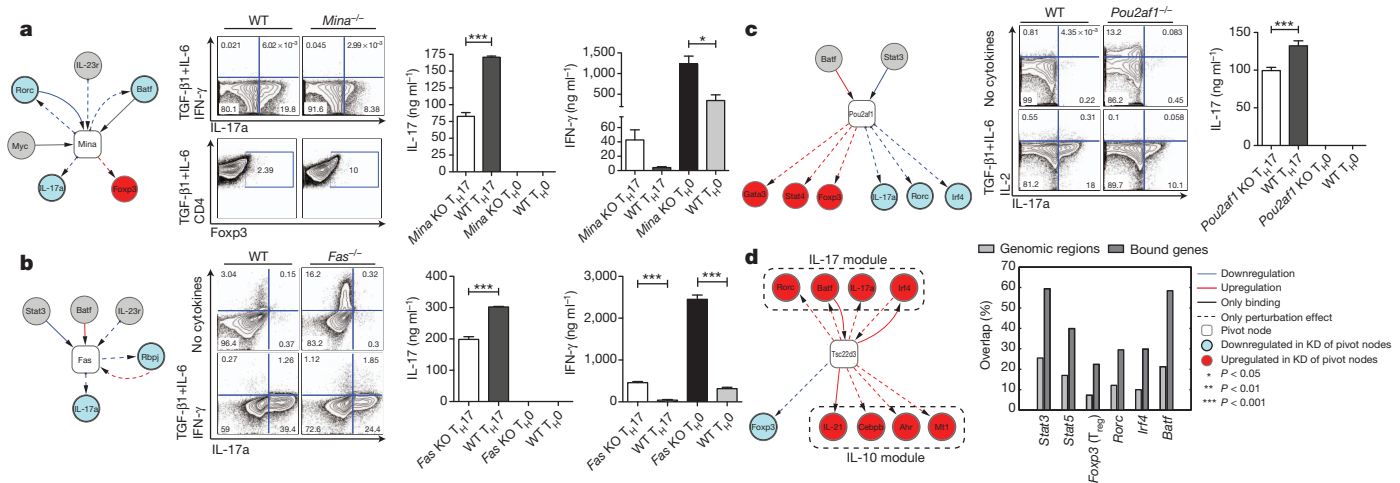


Figure 5 | Mina, Fas, Pou2af1 and Tsc22d3 are key novel regulators affecting the T_H17 differentiation programs. **a–d**, Left: shown are regulatory network models centred on different pivotal regulators (square nodes; **a**, Mina; **b**, Fas; **c**, Pou2af1; **d**, Tsc22d3). In each network, shown are the targets and regulators (round nodes) connected to the pivotal nodes based on perturbation (red and blue dashed edges), transcription factor binding (black solid edges), or both (red and blue solid edges). Genes affected by perturbing the pivotal nodes are coloured (blue, target is downregulated by knockdown of pivotal node; red, target is upregulated). Middle and right panels of **a–c**: intracellular staining and

To analyse the role of Mina further, we measured IL-17a and *Foxp3* expression after differentiation of naive T cells from *Mina*^{-/-} mice. *Mina*^{-/-} cells had decreased IL-17a and increased *Foxp3*⁺ T cells compared to wild-type cells, as detected by intracellular staining (Fig. 5a). Cytokine analysis of the corresponding supernatants confirmed a decrease in IL-17a production and an increase in IFN-γ (Fig. 5a) and TNF-α (Supplementary Fig. 12a). This is consistent with a model where Mina, induced by *ROR-γt* and *Batf*, promotes transcription of *Rorc*, while suppressing induction of *Foxp3*, thus affecting the reciprocal T_{reg}/T_H17 balance³³ by favouring rapid T_H17 differentiation.

Fas promotes the T_H17 program and suppresses IFN-γ

Fas, the TNF receptor superfamily member 6, is another T_H17-positive regulator (Fig. 5b). *Fas* is induced early and is a target of *Stat3* and *Batf* in our model. *Fas* knockdown represses the expression of key T_H17 genes (for example, *Il17a*, *Il17f*, *Hif1a*, *Irf4* and *Rbpj*) and of the induced cluster C14, and promotes the expression of T_H1-related genes, including *Ifngr1* and *Klrd1* (CD94; by RNA-seq, Figs 4, 5b, Supplementary Table 7 and Supplementary Fig. 11). *Fas*- and *Fas*-ligand-deficient mice are resistant to the induction of autoimmune encephalomyelitis (EAE)³⁴, but have no defect in IFN-γ or T_H1 responses. The mechanism underlying this phenomenon has not been identified.

To explore this, we differentiated T cells from *Fas*^{-/-} mice (Fig. 5b and Supplementary Fig. 12c). Consistent with our knockdown analysis, expression of IL-17a was strongly repressed and IFN-γ production was strongly increased under both T_H17 and T_H0 polarizing conditions (Fig. 5b). These results suggest that besides being a death receptor, *Fas* may have an important role in controlling the T_H1/T_H17 balance, and *Fas*^{-/-} mice may be resistant to EAE due to lack of T_H17 cells.

Pou2af1 promotes the T_H17 program and suppresses IL-2 expression

Knockdown of Pou2af1 (also called OBF1) strongly decreases the expression of T_H17 signature genes (Fig. 5c) and of intermediate- and late-induced genes (clusters C19 and C20, $P < 10^{-7}$; Supplementary Tables 5 and 7), while increasing the expression of regulators of other CD4⁺ subsets (for example, *Foxp3*, *Stat4*, *Gata3*) and of genes in

cytokine assays by ELISA or cytometric bead assays (CBA) on culture supernatants at 72 h of *in vitro* differentiated cells from respective knockout mice activated *in vitro* with anti-CD3 plus anti-CD28 with or without T_H17 polarizing cytokines (TGF-β1 plus IL-6). **d**, Middle: ChIP-seq of Tsc22d3. Shown is the proportion of overlap in bound genes (dark grey) or bound regions (light grey) between Tsc22d3 and a host of T_H17 canonical factors (x axis). All results are statistically significant ($P < 10^{-6}$; Hypergeometric score (gene overlap) and Binomial score (region overlap); Supplementary Information).

non-induced clusters (clusters C2 and C16, $P < 10^{-9}$; Supplementary Table 5 and 7). The role of Pou2af1 in T-cell differentiation has not been explored³⁵.

To investigate its effects, we differentiated T cells from *Pou2af1*^{-/-} mice (Fig. 5c and Supplementary Fig. 12b). Compared to wild-type cells, IL-17a production was strongly repressed. Interestingly, IL-2 production was strongly increased in *Pou2af1*^{-/-} T cells under non-polarizing (T_H0) conditions. Thus, Pou2af1 may promote T_H17 differentiation by blocking production of IL-2, a known endogenous repressor of T_H17 cells³⁶. Pou2af1 acts as a transcriptional co-activator of the transcription factors Oct1 or Oct2 (ref. 35). IL-17a production was also strongly repressed in Oct1-deficient cells (Supplementary Fig. 12d), suggesting that Pou2af1 may exert some of its effects through this co-factor.

Tsc22d3 may limit T_H17 generation and inflammation

Knockdown of the TSC22 domain family protein 3 (Tsc22d3) increases the expression of T_H17 cytokines (*Il17a*, *Il21*) and transcription factors (*Rorc*, *Rbpj*, *Batf*), and reduces *Foxp3* expression. Previous studies in macrophages have shown that *Tsc22d3* expression is stimulated by glucocorticoids and IL-10, and it has a key role in their anti-inflammatory and immunosuppressive effects³⁷. Tsc22d3 knockdown in T_H17 cells increased the expression of *Il10* and other key genes that enhance its production (Fig. 5d). Although IL-10 production has been shown^{33,38,39} to render T_H17 cells less pathogenic in autoimmunity, co-production of IL-10 and IL-17a may be the indicated response for clearing certain infections such as *Staphylococcus aureus* at mucosal sites⁴⁰. This suggests a model where Tsc22d3 is part of a negative feedback loop for the induction of a T_H17 cell subtype that co-produces IL-17 and IL-10 and limits their pro-inflammatory capacity. Tsc22d3 is induced in other cells in response to the steroid dexamethasone⁴¹, which represses T_H17 differentiation and *Rorc* expression⁴². Thus, Tsc22d3 may mediate this effect of steroids.

To characterize the role of Tsc22d3 further, we used ChIP-seq to measure its DNA-binding profile in T_H17 cells and RNA-seq following its knockdown to measure its functional effects. There is a significant overlap between Tsc22d3's functional and physical targets ($P < 0.01$, for example, *Il21*, *Irf4*; Supplementary Information and Supplementary Table 8). For example, Tsc22d3 binds to *Il21* and *Irf4*, which also become upregulated in the Tsc22d3 knockdown. Furthermore, the Tsc22d3 binding sites significantly overlap those of major T_H17 factors, including Batf, Stat3, Irf4 and ROR- γ t (>5-fold enrichment; Fig. 5d, Supplementary Table 8 and Supplementary Methods). This suggests a model where Tsc22d3 exerts its T_H17-negative function as a transcriptional repressor that competes with T_H17-positive regulators over binding sites, analogous to previous findings in CD4⁺ regulation^{29,43}.

Discussion

We combined a high-resolution transcriptional time course, novel methods to reconstruct regulatory networks, and innovative nanotechnology to perturb T cells, to construct and validate a network model for T_H17 differentiation. The model consists of three consecutive, densely intra-connected networks, implicates 71 regulators (46 novel), and suggests substantial rewiring in 3 phases. The 71 regulators significantly overlap with genes genetically associated with inflammatory bowel disease⁴⁴ (11 of 71, $P < 10^{-9}$). Building on this model, we systematically ranked 127 putative regulators (82 novel; Supplementary Table 4) and tested top ranking ones experimentally.

We found that the T_H17 regulators are organized into two tightly coupled, self-reinforcing but mutually antagonistic modules, the coordinated action of which may explain how the balance between T_H17, T_{reg} and other effector T-cell subsets is maintained, and how progressive directional differentiation of T_H17 cells is achieved. Within the two modules are 12 novel factors (Figs 4 and 5), which

we further characterized, highlighting four of the factors (others are in Supplementary Note and Supplementary Fig. 13).

A recent study²⁹ systematically ranked T_H17 regulators based on ChIP-seq data for known key factors and transcriptional profiles in wild-type and knockout cells. Whereas their network centred on known core T_H17 transcription factors, our complementary approach perturbed many genes in a physiologically meaningful setting. Reassuringly, their core T_H17 network significantly overlaps with our computationally inferred model (Supplementary Fig. 14).

The wiring of the positive and negative modules (Figs 4 and 5) uncovers some of the functional logic of the T_H17 program, but probably involves both direct and indirect interactions. Our functional model provides an excellent starting point for deciphering the underlying physical interactions with DNA binding profiles³⁰ or protein-protein interactions (accompanying paper⁴⁵). The regulators that we identified are compelling new targets for regulating the T_H17/T_{reg} balance and for switching pathogenic T_H17 into non-pathogenic ones.

METHODS SUMMARY

We measured gene expression profiles at 18 time points (0.5 to 72 h) under T_H17 conditions (IL-6, TGF- β 1) or control (T_H0) using Affymetrix microarrays HT_MG-430A. We detected differentially expressed genes using a consensus over four inference methods, and clustered the genes using *k*-means clustering, with an automatically derived *k*. Temporal regulatory interactions were inferred by looking for significant ($P < 5 \times 10^{-5}$ and fold enrichment >1.5) overlaps between the regulator's putative targets (for example, based on ChIP-seq) and the target gene's cluster (using four clustering schemes). Candidates for perturbation were ordered lexicographically using network-based and expression-based features. Perturbations were done using SiNW for siRNA delivery.

Full Methods and any associated references are available in the online version of the paper.

Received 28 September 2012; accepted 5 February 2013.

Published online 6 March 2013.

- Bettelli, E., Oukka, M. & Kuchroo, V. K. T_H17 cells in the circle of immunity and autoimmunity. *Nature Immunol.* **8**, 345–350 (2007).
- Zhou, L. et al. TGF- β -induced Foxp3 inhibits T_H17 cell differentiation by antagonizing ROR γ t function. *Nature* **453**, 236–240 (2008).
- O'Shea, J. et al. Signal transduction and Th17 cell differentiation. *Microbes Infect.* **11**, 599–611 (2009).
- Zhou, L. & Littman, D. Transcriptional regulatory networks in Th17 cell differentiation. *Curr. Opin. Immunol.* **21**, 146–152 (2009).
- Korn, T., Bettelli, E., Oukka, M. & Kuchroo, V. K. IL-17 and Th17 cells. *Annu. Rev. Immunol.* **27**, 485–517 (2009).
- Amit, I. et al. Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326**, 257–263 (2009).
- Novershtern, N. et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296–309 (2011).
- Litvak, V. et al. Function of C/EBP δ in a regulatory circuit that discriminates between transient and persistent TLR4-induced signals. *Nature Immunol.* **10**, 437–443 (2009).
- Suzuki, H. et al. The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nature Genet.* **41**, 553–562 (2009).
- Shalek, A. K. et al. Vertical silicon nanowires as a universal platform for delivering biomolecules into living cells. *Proc. Natl Acad. Sci. USA* **107**, 1870–1875 (2010).
- Lee, Y. et al. Induction and molecular signature of pathogenic T_H17 cells. *Nature Immunol.* **13**, 991–999 (2012).
- Linhart, C., Halperin, Y. & Shamir, R. Transcription factor and microRNA motif discovery: the Amadeus platform and a compendium of metazoan target sets. *Genome Res.* **18**, 1180–1189 (2008).
- Zheng, G. et al. ITPP: an integrated platform of mammalian transcription factors. *Bioinformatics* **24**, 2416–2417 (2008).
- Wilson, N. K. et al. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* **7**, 532–544 (2010).
- Lachmann, A. et al. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438–2444 (2010).
- Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
- Jiang, C., Xuan, Z., Zhao, F. & Zhang, M. TRED: a transcriptional regulatory element database, new entries and other development. *Nucleic Acids Res.* **35**, D137–D140 (2007).
- Elkon, R., Linhart, C., Sharan, R., Shamir, R. & Shilo, Y. Genome-wide *in silico* identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Res.* **13**, 773–780 (2003).

19. Heng, T. S. & Painter, M. W. The Immunological Genome Project: networks of gene expression in immune cells. *Nature Immunol.* **9**, 1091–1094 (2008).
20. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
21. Dardalhon, V. *et al.* Lentivirus-mediated gene transfer in primary T cells is enhanced by a central DNA flap. *Gene Ther.* **8**, 190–198 (2001).
22. McManus, M. *et al.* Small interfering RNA-mediated gene silencing in T lymphocytes. *J. Immunol.* **169**, 5754 (2002).
23. Shalek, A. K. *et al.* Nanowire-mediated delivery enables functional interrogation of primary immune cells: application to the analysis of chronic lymphocytic leukemia. *Nano Lett.* **12**, 6498–6504 (2012).
24. Segrè, D., Deluna, A., Church, G. M. & Kishony, R. Modular epistasis in yeast metabolism. *Nature Genet.* **37**, 77–83 (2005).
25. Peleg, T., Yosef, N., Rupp, E. & Sharan, R. Network-free inference of knockout effects in yeast. *PLOS Comput. Biol.* **6**, e1000635 (2010).
26. Marson, A. *et al.* Foxp3 occupancy and regulation of key target genes during T-cell stimulation. *Nature* **445**, 931–935 (2007).
27. Zheng, Y. *et al.* Genome-wide analysis of Foxp3 target genes in developing and mature regulatory T cells. *Nature* **445**, 936–940 (2007).
28. Glasmacher, E. *et al.* A genomic regulatory element that directs assembly and function of immune-specific AP-1-IRF complexes. *Science* **338**, 975–980 (2012).
29. Ciofani, M. *et al.* A validated regulatory network for Th17 cell specification. *Cell* **151**, 289–303 (2012).
30. Glasmacher, E. *et al.* A genomic regulatory element that directs assembly and function of immune-specific AP-1-IRF complexes. *Science* **338**, 975–980 (2012).
31. Okamoto, M. *et al.* Mina, an I κ B repressor, controls T helper type 2 bias. *Nature Immunol.* **10**, 872–879 (2009).
32. Hill, J. A. *et al.* Foxp3 transcription-factor-dependent and -independent regulation of the regulatory T cell transcriptional signature. *Immunity* **27**, 786–800 (2007).
33. Korn, T. *et al.* IL-21 initiates an alternative pathway to induce proinflammatory T_H17 cells. *Nature* **448**, 484–487 (2007).
34. Waldner, H., Sobel, R. A., Howard, E. & Kuchroo, V. K. Fas- and FasL-deficient mice are resistant to induction of autoimmune encephalomyelitis. *J. Immunol.* **159**, 3100–3103 (1997).
35. Teitell, M. A. OCA-B regulation of B-cell development and function. *Trends Immunol.* **24**, 546–553 (2003).
36. Laurence, A. *et al.* Interleukin-2 signaling via STAT5 constrains T helper 17 cell generation. *Immunity* **26**, 371–381 (2007).
37. Choi, S.-J. *et al.* Tsc-22 enhances TGF- β signaling by associating with Smad4 and induces erythroid cell differentiation. *Mol. Cell. Biochem.* **271**, 23–28 (2005).
38. Peters, A., Lee, Y. & Kuchroo, V. K. The many faces of Th17 cells. *Curr. Opin. Immunol.* **23**, 702–706 (2011).
39. Chaudhry, A. *et al.* Interleukin-10 signaling in regulatory T cells is required for suppression of Th17 cell-mediated inflammation. *Immunity* **34**, 566–578 (2011).
40. Zielinski, C. E. *et al.* Pathogen-induced human T_H17 cells produce IFN- γ or IL-10 and are regulated by IL-1 β . *Nature* **484**, 514–518 (2012).
41. Jing, Y. *et al.* A mechanistic study on the effect of dexamethasone in moderating cell death in Chinese Hamster Ovary cell cultures. *Biotechnol. Prog.* **28**, 490–496 (2012).
42. Hu, S. M., Luo, Y. L., Lai, W. Y. & Chen, P. F. Effects of dexamethasone on intracellular expression of Th17 cytokine interleukin 17 in asthmatic mice [in Chinese]. *Nan Fang Yi Ke Da Xue Xue Bao* **29**, 1185–1188 (2009).
43. Yang, X. P. *et al.* Opposing regulation of the locus encoding IL-17 through direct, reciprocal actions of STAT3 and STAT5. *Nature Immunol.* **12**, 247–254 (2011).
44. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
45. Wu, C. *et al.* Induction of pathogenic T_H17 cells by inducible salt-sensing kinase SGK1. *Nature* <http://dx.doi.org/10.1038/nature11984> (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Gaffney and L. Solomon for artwork, the Broad's Genomics Platform for sequencing, and D. Kozoriz for cell sorting. Work was supported by NHGRI (1P50HG006193-01 to H.P. and A.R.), NIH Pioneer Awards (5DP1OD003893-03 to H.P., DP1OD003958-01 to A.R.), NIH (NS 30843, NS045937, AI073748 and AI45757 to V.K.K.), National MS Society (RG2571 to V.K.K.), HHMI (A.R.), and the Klarman Cell Observatory (A.R.).

Author Contributions N.Y., A.K.S., J.T.G., H.P., V.K.K. and A.R. conceived the study and designed experiments. N.Y. developed computational methods. N.Y., A.K.S. and J.T.G. analysed the data. A.K.S., J.T.G., H.J., Y.L., A.A., C.W., K.K., S.X., M.J., D.G., R.S., D.Y.L. and J.J.T. conducted the experiments. A.S., M.R.P., P.J.R., M.L.C., M.B. and D.T. provided knockout mice. N.Y., A.K.S., J.T.G., V.K.K., H.P. and A.R. wrote the paper with input from all the authors.

Author Information The microarray, RNA-seq and ChIP-seq data sets have been deposited in the Gene Expression Omnibus database under accession numbers GSE43955, GSE43969, GSE43948 and GSE43949. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.P. (Hongkun_Park@harvard.edu), V.K.K. (vkuchroo@rics.bwh.harvard.edu) or A.R. (aregev@broad.mit.edu).

METHODS

Mice. C57BL/6 wild-type, *Irf1*^{-/-}, *Fas*^{-/-}, *Irf4*^{fl/fl} and *Cd4*^{Cre} mice were obtained from Jackson Laboratory. *Stat1*^{-/-} and 129/Sv control mice were purchased from Taconic. *Il12rb1*^{-/-} mice were provided by P. Kalipada. *Il17ra*^{-/-} mice were provided by J. Kolls. *Irf8*^{fl/fl} mice were provided by K. Ozato. Both *Irf4*^{fl/fl} and *Irf8*^{fl/fl} mice were crossed to *Cd4*^{Cre} mice to generate *Cd4*^{Cre}×*Irf4*^{fl/fl} and *Cd4*^{Cre}×*Irf8*^{fl/fl} mice. All animals were housed and maintained in a conventional pathogen-free facility at the Harvard Institute of Medicine in Boston (IUCAC protocols: 0311-031-14 (V.K.K.) and 0609-058015 (A.R.)). All experiments were performed in accordance to the guidelines outlined by the Harvard Medical Area Standing Committee on Animals at the Harvard Medical School. In addition, spleens from *Mina*^{-/-} mice were provided by M. Bix (IACUC protocol: 453). *Pou2af1*^{-/-} mice were obtained from the laboratory of R. Roeder⁴⁶. Wild-type and *Oclt*^{-/-} fetal livers were obtained at day E12.5 and transplanted into sub-lethally irradiated *Rag1*^{-/-} mice as previously described⁴⁷ (IACUC protocol: 11-09003).

Cell sorting and *in vitro* T-cell differentiation. CD4⁺ T cells were purified from spleen and lymph nodes using anti-CD4 microbeads (Miltenyi Biotec) then stained in PBS with 1% FCS for 20 min at room temperature with anti-CD4-PerCP, anti-CD62L-APC and anti-CD44-PE antibodies (all Biolegend). Naive CD4⁺CD62L^{high}CD44^{low} T cells were sorted using the BD FACSARIA cell sorter. Sorted cells were activated with plate-bound anti-CD3 (2 µg ml⁻¹) and anti-CD28 (2 µg ml⁻¹) in the presence of cytokines. For T_H17 differentiation: 2 ng ml⁻¹ rhTGF-β1 (Miltenyi Biotec), 25 ng ml⁻¹ rmlL-6 (Miltenyi Biotec), 20 ng ml⁻¹ rmlL-23 (Miltenyi Biotec), and 20 ng ml⁻¹ rmlL-β1 (Miltenyi Biotec). Cells were cultured for 0.5–72 h and collected for RNA, intracellular cytokine staining, and flow cytometry.

Flow cytometry and intracellular cytokine staining. Sorted naive T cells were stimulated with phorbol 12-myristate 13-acetate (PMA) (50 ng ml⁻¹, Sigma-aldrich), ionomycin (1 µg ml⁻¹, Sigma-aldrich) and a protein transport inhibitor containing monensin (GolgiStop) (BD Biosciences) for 4 h before detection by staining with antibodies. Surface markers were stained in PBS with 1% FCS for 20 min at room temperature, then subsequently the cells were fixed in Cytoperm/Cytofix (BD Biosciences), permeabilized with Perm/Wash Buffer (BD Biosciences) and stained with Biolegend conjugated antibodies, that is, Brilliant violet 650 anti-mouse IFN-γ (XMG1.2) and allophycocyanin-anti-IL-17A (TC11-18H10.1), diluted in Perm/Wash buffer as described⁴⁸ (Fig. 5 and Supplementary Fig. 11). To measure the time course of ROR-γt protein expression, a phycoerythrin-conjugated anti-retinoid-related orphan receptor-γ was used (B2D), also from eBioscience (Supplementary Fig. 4). Foxp3 staining for cells from knockout mice was performed with the Foxp3 staining kit by eBioscience (00-5523-00) in accordance with their 'One-step protocol for intracellular (nuclear) proteins'. Data were collected using either a FACS Calibur or LSR II (Both BD Biosciences), then analysed using Flow Jo software (Treestar)^{49,50}.

Quantification of cytokine secretion using ELISA. Naive T cells from knockout mice and their wild-type controls were cultured as described above, their supernatants were collected after 72 h, and cytokine concentrations were determined by ELISA (antibodies for IL-17 and IL-10 from BD Bioscience) or by cytometric bead array for the indicated cytokines (BD Bioscience), according to the manufacturers' instructions (Fig. 5 and Supplementary Fig. 11).

Microarray data. Naive T cells were isolated from wild-type mice, and treated with IL-6 and TGF-β1. Affymetrix microarrays HT_MG-430A were used to measure the resulting mRNA levels at 18 different time points (0.5–72 h; Fig. 1b). Cells treated initially with IL-6, TGF-β1 and with addition of IL-23 after 48 h were profiled at four time points (50–72 h). As control, we used time- and culture-matched wild-type naive T cells stimulated under T_H0 conditions. Biological replicates were measured in 8 of the 18 time points (1 h, 4 h, 10 h, 20 h, 30 h, 42 h, 52 h, 60 h) with high reproducibility ($r^2 > 0.98$). For further validation we compared the differentiation time course to published microarray data of T_H17 cells and naive T cells⁵¹ (Supplementary Fig. 1c). In an additional data set, naive T cells were isolated from wild-type and *Il23r*^{-/-} mice, and treated with IL-6, TGF-β1 and IL-23 and profiled at four different time points (49 h, 54 h, 65 h, 72 h). Expression data were pre-processed using the RMA algorithm followed by quantile normalization⁵².

Detecting differentially expressed genes. Differentially expressed genes (comparing to the T_H0 control) were found using four methods: (1) Fold change. Requiring a twofold change (up or down) during at least two time points. (2) Polynomial fit. We used the EDGE software^{53,54}, designed to identify differential expression in time course data, with a threshold of q -value ≤ 0.01 . (3) Sigmoidal fit. We used an algorithm similar to EDGE while replacing the polynomials with a sigmoid function, which is often more adequate for modelling time course gene expression data⁵⁵. We used a threshold of q -value ≤ 0.01 . (4) ANOVA. Gene expression is modelled by: time (using only time points for which we have more

than one replicate) and treatment ('TGF-β1 + IL-6' or 'T_H0'). The model takes into account each variable independently, as well as their interaction. We report cases in which the P value assigned with the treatment parameter or the interaction parameter passed an FDR threshold of 0.01.

Overall, we saw substantial overlap between the methods (average of 82% between any pair of methods). We define the differential expression score of a gene as the number of tests that detected it. As differentially expressed genes, we report cases with differential expression score > 2 .

For the *Il23r*^{-/-} time course (compared to the wild-type T cells) we used methods (1)–(3) (above). Here we used a fold change cutoff of 1.5, and report genes detected by at least two tests.

Clustering. We considered several ways for grouping the differentially expressed genes, based on their time course expression data: (1) for each time point, we defined two groups ((a) all the genes that are overexpressed, and (b) all the genes that are under-expressed relative to T_H0 cells (see below)); (2) for each time point, we defined two groups ((a) all the genes that are induced, and (b) all the genes that are repressed, comparing to the previous time point); (3) k -means clustering using only the T_H17 polarizing conditions. We used the minimal k , such that the within-cluster similarity (average Pearson correlation with the cluster's centroid) was higher than 0.75 for all clusters; and, (4) k -means clustering using a concatenation of the T_H0 and T_H17 profiles.

For methods (1) and (2), to decide whether to include a gene, we considered its original mRNA expression profiles (T_H0, T_H17) and their approximations as sigmoidal functions⁵⁵ (thus filtering transient fluctuations). We require that the fold change levels (compared to T_H0 (method 1) or to the previous time point (method 2)) pass a cutoff defined as the minimum of the following three values: (1) 1.7; (2) mean + s.d. of the histogram of fold changes across all time points; or (3) the maximum fold change across all time points. The clusters presented in Fig. 1b were obtained with method (4). The groupings from methods (1), (2) and (4) are provided in Supplementary Table 2.

Regulatory network inference. We identified potential regulators of T_H17 differentiation by computing overlaps between their putative targets and sets of differentially expressed genes grouped according to methods (1)–(4) above. We assembled regulator–target associations from several sources: (1) *in vivo* DNA binding profiles (typically measured in other cells) of 298 transcriptional regulators^{12–17}; (2) transcriptional responses to the knockout of 11 regulatory proteins^{6,43,49,56–60}; (3) additional potential interactions obtained by applying the Ontogenet algorithm (V. Jojic *et al.*, submitted; regulatory model available at: <http://www.immgen.org/ModsRegs/modules.html>) to data from the mouse ImmGen consortium (<http://www.immgen.org>; January 2010 release¹⁹), which includes 484 microarray samples from 159 cell subsets from the innate and adaptive immune system of mice; (4) a statistical analysis of *cis*-regulatory element enrichment in promoter regions¹⁸; and (5) the transcription factor enrichment module of the IPA software (<http://www.ingenuity.com/>). For every transcription factor in our database, we computed the statistical significance of the overlap between its putative targets and each of the groups defined above using a Fisher's exact test. We include cases where $P < 5 \times 10^{-5}$ and the fold enrichment > 1.5 .

Each edge in the regulatory network was assigned a time stamp based on the expression profiles of its respective regulator and target nodes. For the target node, we considered the time points at which a gene was either differentially expressed or significantly induced or repressed with respect to the previous time point (similarly to grouping methods (1) and (2) above). We defined a regulator node as 'absent' at a given time point if: (i) it was under expressed compared to T_H0; or (ii) the expression is low ($< 20\%$ of the maximum value in time) and the gene was not overexpressed compared to T_H0; or, (iii) up to this point in time the gene was not expressed above a minimal expression value of 100. As an additional constraint, we estimated protein expression levels using the model from ref. 20 and using a sigmoidal fit⁵⁵ for a continuous representation of the temporal expression profiles, and the ProtParam software⁶¹ for estimating protein half-lives. We require that, in a given time point, the predicted protein level be no less than 1.7-fold below the maximum value attained during the time course, and not be less than 1.7-fold below the T_H0 levels. The timing assigned to edges inferred based on a time-point-specific grouping (grouping methods (1) and (2) above) was limited to that specific time point. For instance, if an edge was inferred based on enrichment in the set of genes induced at 1 h (grouping method (2)), it will be assigned a '1 h' time stamp. This same edge could then only have additional time stamps if it was revealed by additional tests.

Selection of nanostring signature genes. The selection of the 275-gene signature (Supplementary Tables 5 and 6) combined several criteria to reflect as many aspects of the differentiation program as was possible. We defined the following requirements: (1) the signature must include all of the transcription factors that belong to a T_H17 microarray signature (comparing to other CD4⁺ T cells⁵¹, see

Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus

Hua-Xin Liao^{1,2*}, Rebecca Lynch^{3*}, Tongqing Zhou^{3*}, Feng Gao^{1,2*}, S. Munir Alam^{1,2}, Scott D. Boyd⁴, Andrew Z. Fire⁴, Krishna M. Roskin⁴, Chaim A. Schramm⁵, Zhenhai Zhang⁵, Jiang Zhu³, Lawrence Shapiro^{3,5}, NISC Comparative Sequencing Program†, James C. Mullikin^{6,7}, S. Gnanakaran⁸, Peter Hraber⁸, Kevin Wiehe^{1,2}, Garnett Kelsoe^{1,2}, Guang Yang^{1,2}, Shi-Mao Xia^{1,2}, David C. Montefiori^{1,2}, Robert Parks^{1,2}, Krissey E. Lloyd^{1,2}, Richard M. Searce^{1,2}, Kelly A. Soderberg^{1,2}, Myron Cohen⁹, Gift Kamanga¹⁰, Mark K. Louder³, Lillian M. Tran³, Yue Chen^{1,2}, Fangping Cai^{1,2}, Sheri Chen^{1,2}, Stephanie Moquin³, Xiulian Du³, M. Gordon Joyce³, Sanjay Srivatsan³, Baoshan Zhang³, Anqi Zheng³, George M. Shaw¹¹, Beatrice H. Hahn¹¹, Thomas B. Kepler¹², Bette T. M. Korber⁸, Peter D. Kwong³, John R. Mascola³ & Barton F. Haynes^{1,2}

Current human immunodeficiency virus-1 (HIV-1) vaccines elicit strain-specific neutralizing antibodies. However, cross-reactive neutralizing antibodies arise in approximately 20% of HIV-1-infected individuals, and details of their generation could provide a blueprint for effective vaccination. Here we report the isolation, evolution and structure of a broadly neutralizing antibody from an African donor followed from the time of infection. The mature antibody, CH103, neutralized approximately 55% of HIV-1 isolates, and its co-crystal structure with the HIV-1 envelope protein gp120 revealed a new loop-based mechanism of CD4-binding-site recognition. Virus and antibody gene sequencing revealed concomitant virus evolution and antibody maturation. Notably, the unmutated common ancestor of the CH103 lineage avidly bound the transmitted/founder HIV-1 envelope glycoprotein, and evolution of antibody neutralization breadth was preceded by extensive viral diversification in and near the CH103 epitope. These data determine the viral and antibody evolution leading to induction of a lineage of HIV-1 broadly neutralizing antibodies, and provide insights into strategies to elicit similar antibodies by vaccination.

Induction of HIV-1 envelope (Env) broadly neutralizing antibodies (BnAbs) is a key goal of HIV-1 vaccine development. BnAbs can target conserved regions that include conformational glycans, the gp41 membrane proximal region, the V1/V2 region, glycan-associated C3/V3 on gp120, and the CD4-binding site^{1–9}. Most mature BnAbs have one or more unusual features (long third complementarity-determining region of the heavy chain (HCDR), polyreactivity for non-HIV-1 antigens, and high levels of somatic mutations), suggesting substantial barriers to their elicitation^{4,10–13}. In particular, CD4-binding site BnAbs have extremely high levels of somatic mutation, suggesting complex or prolonged maturation pathways^{4–7}. Moreover, it has been difficult to find Env proteins that bind with high affinity to BnAb germline or unmutated common ancestors (UCAs), a trait that would be desirable for candidate immunogens for induction of BnAbs^{7,14–18}. Although it has been shown that Env proteins bind to UCAs of BnAbs targeting the gp41 membrane proximal region^{16,19}, and to UCAs of some V1/V2 BnAbs²⁰, so far, heterologous Env proteins have not been identified that bind the UCAs of CD4-binding site BnAb lineages^{7,18,21–23}, although they should exist²¹.

Eighty per cent of heterosexual HIV-1 infections are established by one transmitted/founder virus²⁴. The initial neutralizing antibody response to this virus arises approximately 3 months after transmission and is strain-specific^{25,26}. The antibody response to the transmitted/founder virus drives viral escape, such that virus mutants become

resistant to neutralization by autologous plasma^{25,26}. This antibody-virus race leads to poor or restricted specificities of neutralizing antibodies in ~80% of patients; however in ~20% of patients, evolved variants of the transmitted/founder virus induce antibodies with considerable neutralization breadth, such as BnAbs^{2,20,27–33}.

There are several potential molecular routes by which antibodies to HIV-1 may evolve, and indeed, types of antibody with different neutralizing specificities may follow different routes^{6,11,15,34}. Because the initial autologous neutralizing antibody response is specific for the transmitted/founder virus³¹, some transmitted/founder Env proteins might be predisposed to binding the germ line or UCA of the observed BnAb in those rare patients that make BnAbs. Thus, although neutralizing breadth generally is not observed until chronic infection, a precise understanding of the interaction between virus evolution and maturing BnAb lineages in early infection may provide insight into events that ultimately lead to BnAb development. BnAbs studied so far have only been isolated from individuals who were sampled during chronic infection^{1,3–7,20,27,29}. Thus, the evolutionary trajectories of virus and antibody from the time of virus transmission to the development of broad neutralization remain unknown.

We and others have proposed vaccine strategies that begin by targeting UCAs, the putative naive B-cell receptors of BnAbs with relevant Env immunogens to trigger antibody lineages with potential ultimately to develop breadth^{6,11,13–16,18,19,21}. This would be followed by vaccination

¹Duke University Human Vaccine Institute, Departments of Medicine and Immunology, Duke University School of Medicine, Durham, North Carolina 27710, USA. ²Duke Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery, Durham, North Carolina 27710, USA. ³Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. ⁴Department of Pathology, Stanford University, Palo Alto, California 94305, USA. ⁵Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, USA. ⁶NISC Comparative Sequencing Program, NIH, Bethesda, Maryland 20892, USA. ⁷NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. ⁸Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87544, USA. ⁹Departments of Medicine, Epidemiology and Microbiology and Immunology, University of North Carolina, North Carolina 27599, USA. ¹⁰University of North Carolina Project, Kamuzu Central Hospital, Lilongwe, Malawi. ¹¹Departments of Medicine and Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ¹²Department of Microbiology, Boston University, Boston, Massachusetts 02215, USA.

*These authors contributed equally to this work.

†A full list of participants and their affiliations appears at the end of the paper.

with Env proteins specifically selected to stimulate somatic mutation pathways that give rise to BnAbs. Both aspects of this strategy have proved challenging owing to a lack of knowledge of specific Env proteins capable of interacting with UCAs and early intermediate antibodies of BnAbs.

Here we report the isolation of the CH103 CD4-binding site BnAb clonal lineage from an African patient, CH505, who was followed from acute HIV-1 infection to BnAb development. We show that the CH103 BnAb lineage is less mutated than most other CD4-binding site BnAbs, and may be first detectable as early as 14 weeks after HIV-1 infection. Early autologous neutralization by antibodies in this lineage triggered virus escape, but rapid and extensive Env evolution in and near the epitope region preceded the acquisition of plasma antibody neutralization breadth defined as neutralization of heterologous viruses. Analysis of the co-crystal structure of the CH103 Fab fragment and a gp120 core demonstrated a new loop-binding mode of antibody neutralization.

Isolation of the CH103 BnAb lineage

The CH505 donor was enrolled in the CHAVI001 acute HIV-1 infection cohort³⁵ approximately 4 weeks after HIV-1 infection (Supplementary Fig. 1) and followed for more than 3 years. Single genome amplification of 53 plasma viral Env gp160 RNAs²⁴ from 4 weeks after transmission identified a single clade C transmitted/founder virus. Serological analysis demonstrated the development of autologous neutralizing antibodies at 14 weeks, CD4-binding site antibodies that bound to a recombinant Env protein (resurfaced stabilized core 3 (RSC3))⁵ at 53 weeks, and evolution of plasma cross-reactive neutralizing activity from 41–92 weeks after transmission³⁰ (Fig. 1, Supplementary Table 1 and Supplementary Fig. 2). The natural variable regions of heavy-chain (*VH*DJH) and light-chain (*VL*JL) gene pairs of antibodies CH103, CH104 and CH106 were isolated from peripheral blood

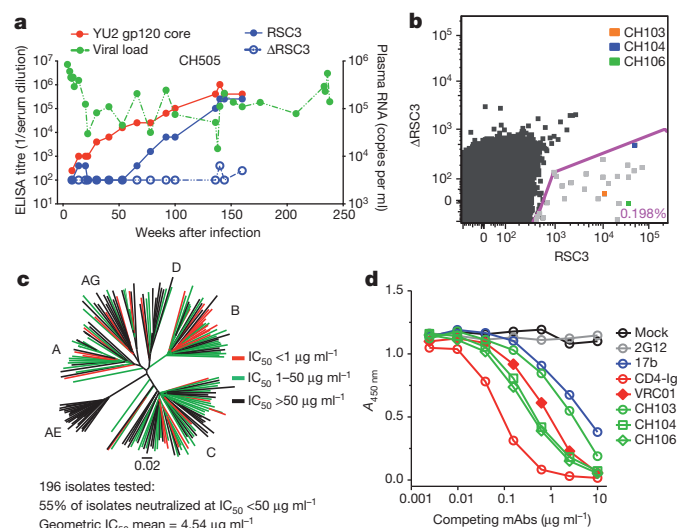


Figure 1 | Development of neutralization breadth in donor CH505 and isolation of antibodies. **a**, Shown are HIV-1 viral RNA copies and reactivity of longitudinal plasmas samples with HIV-1 YU2 gp120 core, RSC3 and negative control RSC3Δ371Ile (ΔRSC3) proteins. **b**, PBMCs from week 136 were used for sorting CD19⁺, CD20⁺, IgG⁺, RSC3⁺ and ΔRSC3⁺ memory B cells (0.198%). Individual cells indicated as orange, blue and green dots yielded monoclonal antibodies CH103, CH104 and CH106, respectively, as identified by index sorting. **c**, The neutralization potency and breadth of the CH103 antibody are displayed using a neighbour-joining tree created with the PHYILIP package. The individual tree branches for 196 HIV-1 Env proteins representing major circulating clades are coloured according to the neutralization IC₅₀ values as indicated. **d**, Cross competition of CH103 binding to YU2 gp120 by the indicated HIV-1 antibodies, and soluble CD4-Ig was determined by ELISA. mAbs, monoclonal antibodies.

mononuclear cells (PBMCs) at 136 weeks after transmission by flow sorting of memory B cells that bound RSC3 Env protein^{5,13,36} (Fig. 1b). The *VH*DJH gene of antibody CH105 was similarly isolated, but no *VL*JL gene was identified from the same cell. Analysis of characteristics of *VH*DJH (VH4–59, posterior probability (PP) = 0.99; D3–16, PP = 0.74; JH4, PP = 1.00) and *VL*JL (VL3–1, PP = 1.00; Jλ1, PP = 1.00) rearrangements in monoclonal antibodies CH103, CH104, CH105 and CH106 demonstrated that these antibodies were representatives of a single clonal lineage that we designated as the CH103 clonal lineage (Fig. 2 and Supplementary Table 2).

Neutralization assays using a previously described^{5,37} panel of 196 geographically and genetically diverse Env-pseudoviruses representing the major circulated genetic subtypes and circulating recombinant forms demonstrated that CH103 neutralized 55% of viral isolates, with a geometric mean half-maximum inhibitory concentration

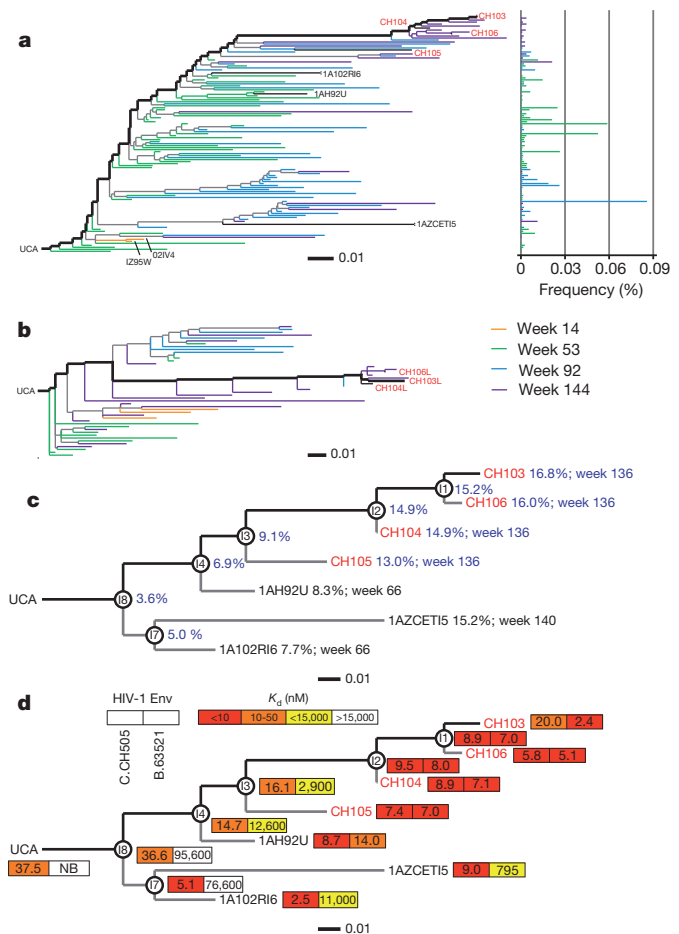


Figure 2 | CH103 clonal family with time of appearance, *VH*DJH mutations and HIV-1 Env reactivity. **a**, **b**, Phylogenies of *VH*DJH (a) and *VL*JL (b) sequences from sorted single memory B cells and pyrosequencing. The ancestral reconstructions for each were performed as described in the Methods. The phylogenetic trees were subsequently computed using neighbour-joining on the complete set of DNA sequences (see Methods) to illustrate the correspondence of sampling date and read abundance in the context of the clonal history. Within time-point *VH* monophyletic clades are collapsed to single branches; variant frequencies are indicated on the right. Isolated mature antibodies are red, pyrosequencing-derived sequences are black. The inferred evolutionary paths to observed matured antibodies are bold. **c**, Maximum-likelihood phylogram showing the CH103 lineage with the inferred intermediates (circles, I1–4, I7 and I8), and percentage mutated *VH* sites and timing (blue), indicated. **d**, Binding affinities (*K_d*, nM) of antibodies to autologous subtype C CH505 (C.CH505; left box) and heterologous B.63521 (right box) were measured by surface plasmon resonance.

(IC₅₀) of 4.54 mg ml⁻¹ among sensitive isolates (Fig. 1c and Supplementary Table 3). Enzyme-linked immunosorbent assay (ELISA) cross-competition analysis demonstrated that CH103 binding to gp120 was competed by known CD4-binding site ligands such as monoclonal antibody VRC01 and the chimeric protein CD4-Ig (Fig. 1d); CH103 binding to RSC3 Env was also substantially diminished by gp120, with Pro363Asn and Δ371Ile mutations known to reduce the binding of most CD4-binding site monoclonal antibodies^{5,30} (Supplementary Fig. 3).

Molecular characterization of the CH103 BnAb lineage

The RSC3 probe isolated CH103, CH104, CH105 and CH106 BnAbs by single-cell flow sorting. The CH103 clonal lineage was enriched by VHDJH and VLJL sequences identified by pyrosequencing PBMC DNA^{34,38} obtained 66 and 140 weeks after transmission, and complementary DNA antibody transcripts⁶ obtained 6, 14, 53, 92 and 144 weeks after transmission. From pyrosequencing of antibody gene transcripts, we found 457 unique heavy- and 171 unique light-chain clonal members (Fig. 2a, b). For comprehensive study, a representative 14-member BnAb pathway was reconstructed from VHDJH sequences (1AH92U, 1AZCET and 1A102R) recovered by pyrosequencing, and VHDJH genes of the inferred intermediate (I) antibodies (I1–I4, I7, I8)^{11,16,34} (T. B. Kepler, manuscript submitted; <http://arxiv.org/abs/1303.0424>) that were paired and expressed with either the UCA or I2 VLJL depending on the genetic distance of the VHDJH to either the UCA or mature antibodies (Fig. 2c and Supplementary Table 2). The mature CH103, CH104 and CH106 antibodies were paired with their natural VLJL. The CH105 natural VHDJH isolated from RSC3 memory B-cell sorting was paired with the VLJL of I2.

Whereas the VHDJH mutation frequencies (calculated as described in the Methods) of the published CD4-binding site BnAbs VRC01, CH31 and NIH45-46 are 30–36% (refs 5–7, 22, 39), the VHDJH frequencies of CH103 lineage CH103, CH104, CH105 and CH106 are 13–17% (Fig. 2c). Furthermore, antibodies in CH103 clonal lineage do not contain the large (>3 nucleotides) insertion or deletion mutations common in the VRC01 class of BnAbs^{1–3}, with the exception of the VLJL of CH103, which contained a three amino-acid light-chain complementarity-determining region 1 (LCDRI) deletion.

It has been proposed that one reason that CD4-binding site BnAbs are difficult to induce is because heterologous HIV-1 Env proteins do not bind their UCAs^{7,18,22}. We wondered, however, whether the CH505 transmitted/founder Env, the initial driving antigen for the CH103 BnAb lineage, would preferentially bind to early CH103 clonal lineage members and the UCA compared to heterologous Env proteins. Indeed, a heterologous gp120 transmitted/founder Env, subtype B 63521 (B.63521), did not bind to the CH103 UCA (Fig. 2d) but did bind to later members of the clonal lineage. Affinity for this heterologous Env protein increased four orders of magnitude during somatic evolution of the CH103 lineage, with maximal dissociation constant (*K_d*) values of 2.4–7.0 nM in the mature CH103–CH106 monoclonal antibodies (Fig. 2d). The CH103 UCA monoclonal antibody did not bind to heterologous transmitted/founder Env proteins AE.427299, B.9021 and C.1086 (Supplementary Table 4), confirming lack of heterologous Env binding to CD4-binding site UCAs. Moreover, the gp120 Env RSC3 protein was also not bound by the CH103 UCA and earlier members of the clonal lineage (Supplementary Fig. 3a), and no binding was seen with RSC3 mutant proteins known to disrupt CD4-binding site BnAb binding (Supplementary Fig. 3b).

In contrast to heterologous Env proteins, the CH505 transmitted/founder Env gp140 bound well to all of the candidate UCAs (Supplementary Table 5), with the highest UCA affinity of *K_d* = 37.5 nM. In addition, the CH505 transmitted/founder Env gp140 was recognized by all members of the CH103 clonal lineage (Fig. 2d). Whereas affinity to the heterologous transmitted/founder Env B.63521 increased by more than four orders of magnitude as the CH103 lineage matured, affinity for the CH505 transmitted/founder Env increased by no more than tenfold (Fig. 2d). To demonstrate Env escape from CH103 lineage

members directly, autologous recombinant gp140 Env proteins isolated at weeks 30, 53 and 78 after infection were expressed and compared with the CH505 transmitted/founder Env for binding to the BnAb arm of the CH103 clonal lineage (Supplementary Table 6 and Supplementary Fig. 4). Escape-mutant Env proteins could be isolated that were progressively less reactive with the CH103 clonal lineage members. Env proteins isolated at weeks 30, 53 and 78 lost UCA reactivity and only bound intermediate antibodies 3, 2 and 1, as well as BnAbs CH103, CH104, CH105 and CH106 (Supplementary Table 6). In addition, two Env escape mutants from week-78 viruses also lost either strong reactivity to all intermediate antibodies or all lineage members (Supplementary Table 6).

To quantify CH103 clonal variants from initial generation to induction of broad and potent neutralization, we used pyrosequencing of antibody cDNA transcripts from five time points, weeks 6, 14, 53, 92 and 144 after transmission (Supplementary Table 7). We found two VHDJH chains closely related to, and possibly members of, the CH103 clonal lineage (Fig. 2a, Supplementary Table 7). Moreover, one of these VHDJH chains when reconstituted in a full IgG1 backbone and expressed with the UCA VLJL weakly bound the CH505 transmitted/founder Env gp140 at an end-point titre of 11 μg ml⁻¹ (Fig. 2a). These reconstructed antibodies were present concomitant with CH505 plasma autologous neutralizing activity at 14 weeks after transmission (Supplementary Fig. 2). Antibodies that bound the CH505 transmitted/founder Env were present in plasma as early as 4 weeks after transmission (data not shown). Both CH103 lineage VHDJH and VLJL sequences peaked at week 53, with 230 and 83 unique transcripts, respectively. VHDJH clonal members fell to 46 at week 144, and VLJL members dropped to 76 at week 144.

Polyreactivity is a common trait of BnAbs, suggesting that the generation of some BnAbs may be controlled by tolerance mechanisms^{10,21,40}. Conversely, polyreactivity can arise during the somatic evolution of B cells in germinal centres as a normal component of B-cell development⁴¹. The CH103 clonal lineage was evaluated for polyreactivity as measured by HEP-2 cell reactivity and binding to a panel of autoantigens¹⁰. Although earlier members of the CH103 clonal lineage were not polyreactive by these measures, polyreactivity was acquired together with BnAb activity by the intermediate antibody I2, I1 and clonal members CH103, CH104 and CH106 (Supplementary Fig. 5a, b). The BnAbs CH106 and intermediate antibody I1 also demonstrated polyreactivity in protein arrays with specific reactivity to several human autoantigens, including elongation factor-2 kinase and ubiquitin-protein ligase E3A (Supplementary Fig. 5c, d).

Structure of CH103 in complex with HIV-1 gp120

Crystals of the complex between the CH103 Fab fragment and the ZM176.66 strain of HIV diffracted to 3.25 Å resolution, and molecular replacement identified solutions for CH103 Fab and for the outer domain of gp120 (Fig. 3a). Inspection of the CH103–gp120 crystal lattice (Supplementary Fig. 6) indicated that the absence of the gp120 inner domain was probably related to proteolytic degradation of the extended gp120 core to an outer domain fragment. Refinement to a *R_{work}*/*R_{free}* ratio of 19.6%/25.6% (Supplementary Table 8) confirmed a lack of electron density for gp120 residues amino-terminal to gp120 residue Val 255 or carboxy-terminal to Gly 472 (gp120 residues are numbered according to standard HXB2 nomenclature), and no electron density was observed for gp120 residues 301–324 (V3), 398–411 (V4) and 421–439 (β20–21). Superposition of the ordered portions of gp120 in complex with CH103 with the fully extended gp120 core bound by antibody VRC01 (ref. 7) indicated a highly similar structure (Cα root mean squared deviation (r.m.s.d.) 1.16 Å) (Fig. 3b). Despite missing portions of core gp120, the entire CH103 epitope seemed to be present in the electron density for the experimentally observed gp120 outer domain.

The surface bound by CH103 formed an elongated patch with dimensions of ~40 × 10 Å, which stretched across the site of initial

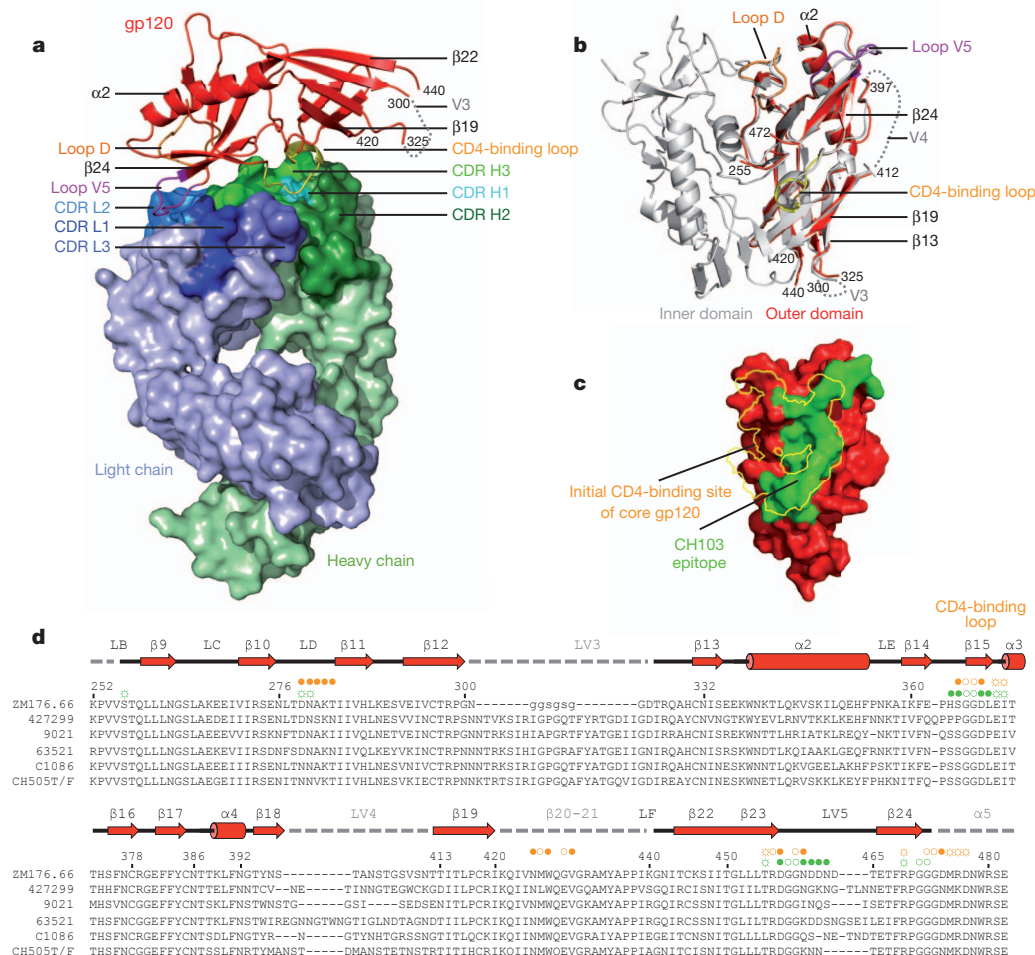


Figure 3 | Structure of antibody CH103 in complex with the outer domain of HIV-1 gp120. **a**, Overall structure of the CH103–gp120 complex, with gp120 polypeptide depicted in red ribbon and CH103 shown as a molecular surface (heavy chain in green, light chain in blue). Major CH103-binding regions on gp120 are coloured orange for loop D, yellow for the CD4-binding site, and purple for loop V5. **b**, Superposition of the outer domain of gp120 bound by CH103 (red), and core gp120 bound by VRC01 (grey), with polypeptide shown in ribbon representation. **c**, CH103 epitope (green) on gp120 outer domain (red), with the initial CD4-binding site superposed (yellow boundaries) in surface representation. **d**, Sequence alignment of outer domains of the crystallized gp120 shown on the first line, and diverse HIV-1 Env proteins recognized by CH103. Secondary structure elements are labelled above the alignment, with grey dashed lines indicating disordered regions. Symbols in yellow or green denote gp120 outer domain contacts for CD4 and CH103, respectively, with open circles representing main-chain contacts, open circles with rays representing side-chain contacts, and filled circles representing both main-chain and side-chain contacts.

CD4 contact on the outer domain of gp120 (Fig. 3c). The gp120 surface recognized by CH103 correlated well with the initial site of CD4 contact; of the residues contacted by CH103, only eight were not predicted to interact with CD4. CH103 interacted with these gp120 residues through side-chain contact with Ser 256 in loop D, main- and side-chain contacts with His 364 and Leu 369 in the CD4-binding loop, and main- and side-chain contacts with Asn 463 and Asp 464 in the V5 loop (Fig. 3d). Notably, residue 463 is a predicted site of *N*-linked glycosylation in strain ZM176.66 as well as in the autologous CH505 virus, but electron density for an *N*-linked glycan was not observed. Overall, of the 22 residues that monoclonal antibody CH103 was observed to contact on gp120, 14 were expected to interact with CD4 (16 of these residues with antibody VRC01), providing a structural basis for the CD4-epitope specificity of CH103 and its broad recognition (Supplementary Table 9).

Residues 1–215 on the antibody heavy chain and 1–209 on the light chain showed well-defined backbone densities. Overall, CH103 uses a CDR H3 dominated mode of interaction, although all six of the complementarity-determining regions (CDRs) interacted with gp120 as well as the light-chain framework region 3 (FWR3) (Supplementary Fig. 7a, b and Supplementary Tables 10 and 11). It is important to note that ~40% of the antibody contact surface was altered by somatic mutation in the HCDR2, LCDR1, LCDR2 and FWR3. In particular, residues 56 on the heavy chain, and residues 50, 51 and 66 on the light chain are altered by somatic mutation to form hydrogen bonds with the CD4-binding loop, loop D and loop V5 of gp120. Nevertheless, 88% of the CH103 V_HD_H and 44% of the V_LJ_L contact areas were with amino acids unmutated in the CH103 germ line, potentially providing an explanation for the robust binding of the transmitted/founder

Env to the CH103 UCA (Supplementary Fig. 7c, d and Supplementary Table 12).

Evolution of transmitted/founder Env sequences

Using single genome amplification and sequencing²⁴ we tracked the evolution of CH505 *env* genes longitudinally from the transmitted/founder virus to 160 weeks after transmission (Fig. 4 and Supplementary Fig. 8). The earliest recurrent mutation in Env, Asn279Lys (HIV-1 HXB2 numbering), was found 4 weeks after infection, and was in Env loop D in a CH103 contact residue. By week 14, additional mutations in loop D appeared, followed by mutations and insertions in the V1 loop at week 20. Insertions and mutations in the V5 loop began to accumulate by week 30 (Fig. 4). Thus, the transmitted/founder virus began to diversify in key CD4 contact regions starting within 3 months of infection (Supplementary Figs 8 and 9). Loop D and V5 mutations were directly in or adjacent to CH103/Env contact residues. Although the V1 region was not included in the CH103–Env co-crystal, the observed V1 CH505 Env mutations were adjacent to contact residues for CD4 and VRC01 so are likely to be relevant. It is also possible that early V1 insertions (Fig. 4) were selected by inhibiting access to the CD4-binding site in the trimer or that they arose in response to early T-cell pressure. CD4-binding-loop mutations were present by week 78. Once regions that could directly affect CH103-lineage binding began to evolve (loop D, V5, the CD4-binding loop, and possibly loop V1), they were under sustained positive selective pressure throughout the study period (Fig. 4, Supplementary Figs 8 and 9 and Supplementary Table 13).

Considerable within-sample virus variability was evident in Env regions that could affect CH103-lineage antibody binding, and

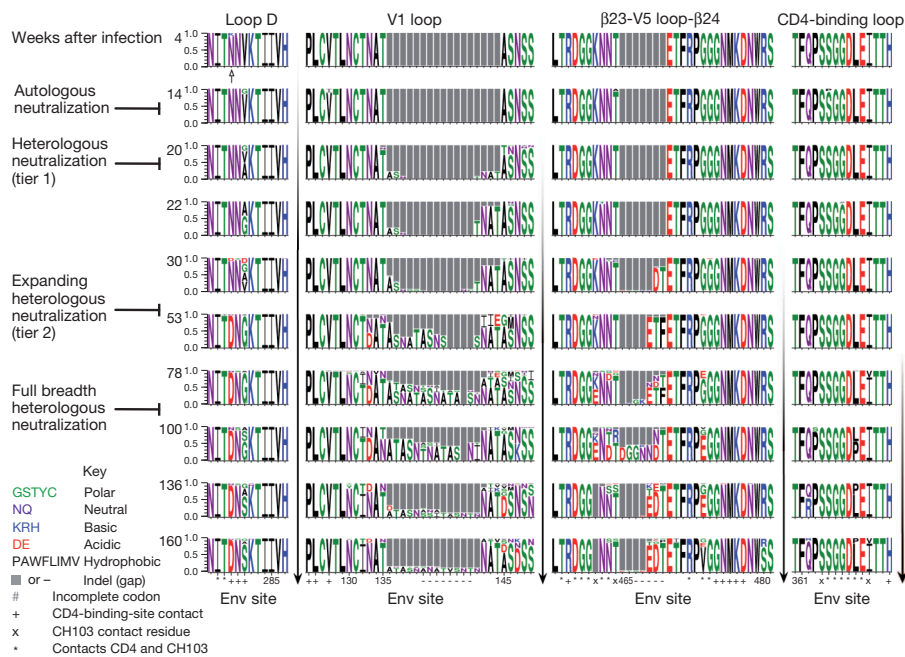


Figure 4 | Sequence logo displaying variation in key regions of CH505 Env proteins. The frequency of each amino acid variant per site is indicated by its height, deletions are indicated by grey bars. The first recurring mutation, Asn279Lys, appears at week 4 (open arrow). The timing of BnAb activity development (from Supplementary Fig. 2 and Supplementary Table 1) is on the left. Viral diversification, which precedes acquisition of breadth, is highlighted by vertical arrows to the right of each region. CD4 and CH103 contact residues, and amino acid position numbers based on HIV-1 HXB2, are shown along the base of each logo column.

diversification in these regions preceded neutralization breadth. Expanding diversification early in viral evolution (4–22 weeks after transmission; Supplementary Figs 8 and 9) coincided with autologous neutralizing antibody development, consistent with autologous neutralizing antibody escape mutations. Mutations that accumulated from weeks 41 to 78 in CH505 Env contact regions immediately preceded development of neutralizing antibody breadth (Fig. 4 and Supplementary Figs 8 and 9). By weeks 30–53, extensive within-sample diversity resulted from both point mutations in and around CH103 contact residues, and to several insertions and deletions in V1 and V5 (Supplementary Fig. 9). A strong selective pressure seems to have come into play between weeks 30 and 53, perhaps due to autologous neutralization escape, and neutralization breadth developed after this point (Fig. 4 and Supplementary Figs 8 and 9). Importantly, owing to apparent strong positive selective pressure between weeks 30 and 53, there was a marked shift in the viral population that is evident in the phylogenetic tree, such that only viruses carrying multiple mutations relative to the transmitted/founder, particularly in CH103 contact regions, persisted after week 30. This was followed by extreme and increasing within time-point diversification in key epitope regions, beginning at week 53 (Supplementary Fig. 9). Emergence of antibodies with neutralization breadth occurred during this time (Supplementary Fig. 2 and Supplementary Table 1). Thus, plasma breadth evolved in the presence of highly diverse forms of the CH103 epitope contact regions (Fig. 4 and Supplementary Fig. 2).

To evaluate and compare the immune pressure on amino acids in the region of CH103 and CD4 contacts, we compared the frequency of mutations in evolving transmitted/founder sequences of patient CH505 during the first year of infection and in 16 other acutely infected subjects followed over time (Supplementary Fig. 10). The accumulation of mutations in the CH505 viral population was concentrated in regions likely to be associated with escape from the CH103 lineage (Supplementary Fig. 10a), and diversification of these regions was far more extensive during the first six months of infection in CH505 than in other subjects (Supplementary Fig. 10b). However, by one year into their infections, viruses from the other subjects had also begun to acquire mutations in these regions. Thus, the early and continuing accumulation of mutations in CH103 contact regions may have potentiated the early development of neutralizing antibody breadth in patient CH505.

Neutralization of viruses and the CH103 lineage

Heterologous BnAb activity was confined to the later members (I3 and later) of the BnAb arm of the CH103 lineage, as manifested by their neutralization capacity of pseudoviruses carrying tier 2 Env proteins A.Q842 and B.BG1168 (Fig. 5a). Similar results were seen with Env proteins A.Q168, B.JRFL, B.SF162 and C.ZM106 (Supplementary Tables 14 and 15). By contrast, neutralizing activity of clonal lineage members against the autologous transmitted/founder Env pseudovirus appeared earlier, with measurable neutralization of the CH505 transmitted/founder virus by all members of the lineage after the UCA except monoclonal antibody 1AH92U (Fig. 5a). Thus, within the CH103 lineage, early intermediate antibodies only neutralized the transmitted/founder virus, whereas later intermediate antibodies gained neutralization breadth, indicating evolution of neutralization breadth with affinity maturation, and CH103–CH106 BnAbs evolved from an early autologous neutralizing antibody response. Moreover, the clonal lineage was heterogeneous, with an arm of the lineage represented in Fig. 5a evolving neutralization breadth and another antibody arm capable of mediating only autologous transmitted/founder virus neutralization. Although some escape-mutant viruses are clearly emerging over time (Supplementary Table 4), it is important to point out that, although the escape-mutant viruses are driving BnAb evolution, the BnAbs remained capable of neutralizing the CH505 transmitted/founder virus (Fig. 5a). Of note, the earliest mutations in the heavy-chain lineage clustered near the contact points with gp120, and these remained fixed throughout the period of study, whereas mutations that accumulated later tended to be further from the binding site and may be affecting binding less directly (Supplementary Fig. 11). Thus, stimulation of the CH103 BnAbs occurs in a manner to retain reactivity with the core CD4-binding site epitope present on the transmitted/founder Env. One possibility that might explain this is that the footprint of UCA binding contracts to the central core binding site of the CH103 mature antibody. Obtaining a crystal structure of the UCA with the transmitted/founder Env should inform this notion. Another possibility is that because affinity maturation is occurring in the presence of highly diverse forms of the CD4-binding site epitope, antibodies that favour tolerance of variation in and near the epitope are selected instead of those antibodies that acquire increased affinity for particular escape Env proteins. In both scenarios, persistence of activity to the transmitted/founder form and early viral variants

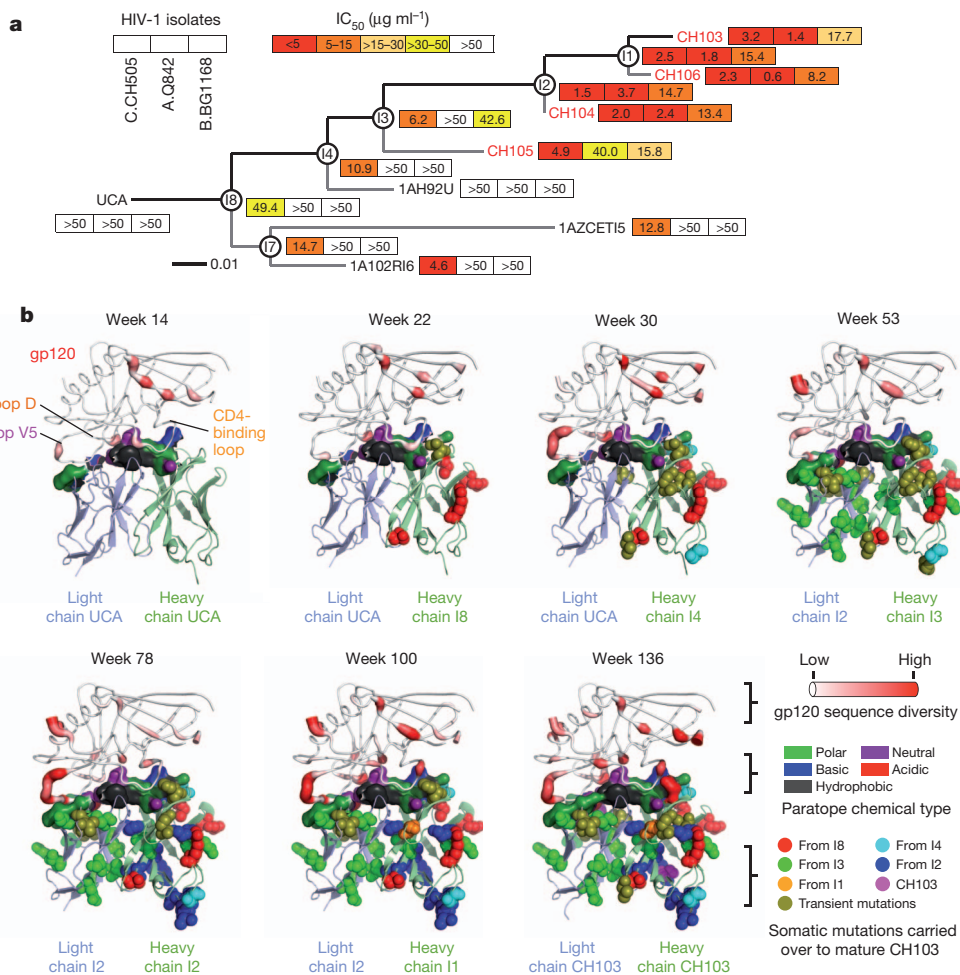


Figure 5 | Development of neutralization breadth in the CH103 clonal lineage.

a, Phylogenetic CH103 clonal lineage tree showing the IC₅₀ (μg ml⁻¹) of neutralization of the autologous transmitted/founder (C.CH505), heterologous tier clades A (A.Q842) and B (B.BG1168) viruses as indicated. **b**, Interaction between evolving virus and developing clonal lineage mapped on to models of CH103 developmental variants and contemporaneous virus. The outer domain of HIV gp120 is depicted in worm representation, with worm thickness and colour (white to red) mapping the degree of per-site sequence diversity at each time point. Models of antibody intermediates are shown in cartoon diagram, with somatic mutations at each time point highlighted in spheres and coloured red for mutations carried over from I8 to mature antibody, cyan for mutations carried over from I3 to mature antibody, green for mutations carried over from I2 to mature antibody, blue for mutations carried over from I1 to mature antibody, and orange for mutations carried over from I1 to mature antibody, and magenta for CH103 mutations from I1. Transient mutations that did not carry all the way to mature antibody are coloured in deep olive. The antibody (paratope) residues are shown in surface representation and coloured by their chemical types as indicated.

would be expected. Figure 5b and Supplementary Fig. 11 show views of accumulations of mutations or entropy during the parallel evolution of the antibody paratope and the Env epitope bound by monoclonal antibody CH103.

Vaccine implications

In this study, we demonstrate that the binding of a transmitted/founder Env to a UCA B-cell receptor of a BnAb lineage was responsible for the induction of broad neutralizing antibodies, thus providing a logical starting place for vaccine-induced CD4-binding site BnAb clonal activation and expansion. Importantly, the number of mutations required to achieve neutralization breadth was reduced in the CH103 lineage compared to most CD4-binding site BnAbs, although the CH103 lineage had reduced neutralization breadth compared to more mutated CD4-binding site BnAbs. Thus, this type of BnAb lineage may be less challenging to attempt to recapitulate by vaccination. By tracking viral evolution through early infection we found that intense selection and epitope diversification in the transmitted/founder virus preceded the acquisition of neutralizing antibody breadth in this individual—thus demonstrating the viral variants associated with development of BnAbs directly from autologous neutralizing antibodies and illuminating a pathway for induction of similar B-cell lineages.

These data have implications for understanding the B-cell maturation pathways of the CH103 lineage and for replicating similar pathways in a vaccine setting. First, we demonstrate in CH505 that BnAbs were driven by sequential Env evolution beginning as early as 14 weeks after transmission, a time period compatible with induction of this type of BnAb lineage with a vaccine given the correct set of immunogens. Second, whereas heterologous Env proteins did not bind with UCAs or early intermediate antibodies of this lineage, the CH505 transmitted/

founder Env bound remarkably well to the CH103 UCA, and subsequent Env proteins bound with increased affinity to later clonal lineage members. This suggests that immunizations with similar sequences of Env or Env subunits may drive similar lineages. Third, the CH103 lineage is less complicated than those of the VRC01 class of antibodies because antibodies in this lineage have fewer somatic mutations, and no indels, except CH103 V1, which has a deletion of three amino acid residues in the LCDRI region. It should also be noted that our study is in one patient. Nonetheless, in each BnAb patient, analysis of viral evolution should determine a similar pathway of evolved Env proteins that induce BnAb breadth. The observation that rhesus macaques infected with the CCR5-tropic simian/human immunodeficiency virus (SHIV)-AD8 virus frequently develop neutralization breadth⁴² suggests that certain Env proteins may be more likely to induce breadth and potency than others.

Polyreactivity to host molecules in the CH103 lineage arose during affinity maturation in the periphery coincident with BnAb activity. This finding is compatible with the hypothesis that BnAbs may be derived from an inherently polyreactive pool of B cells, with polyreactivity providing a neutralization advantage via heterologation of Env and host molecules^{21,43}. Alternatively, as CH103 affinity maturation involves adapting to the simultaneous presence of diverse co-circulating forms of the epitope⁴⁴, the selection of antibodies that can interact with extensive escape-generated epitope diversification may be an evolutionary force that also drives incidental acquisition of polyreactivity.

Thus, a candidate vaccine concept could be to use the CH505 transmitted/founder Env or Env subunits (to avoid dominant Env non-neutralizing epitopes) to initially activate an appropriate naive B-cell response, followed by boosting with subsequently evolved

CH505 Env variants either given in combination, to mimic the high diversity observed *in vivo* during affinity maturation, or in series, using vaccine immunogens specifically selected to trigger the appropriate maturation pathway by high-affinity binding to UCA and antibody intermediates¹¹. These data demonstrate the power of studying subjects followed from the transmission event to the development of plasma BnAb activity for concomitant isolation of both transmitted/founder viruses and their evolved quasiespecies along with the clonal lineage of induced BnAbs. The finding that the transmitted/founder Env can be the stimulator of a potent BnAb and bind optimally to that BnAb UCA is a crucial insight for vaccine design, and could allow the induction of BnAbs by targeting UCAs and intermediate ancestors of BnAb clonal lineage trees¹¹.

METHODS SUMMARY

Serial blood samples were collected from an HIV-1-infected subject CH505 from 4 to 236 weeks after infection. Monoclonal antibodies CH103, CH104 and CH106 were generated by the isolation, amplification and cloning of single RSC3-specific memory B cells as described^{5–7,22,36}. *VH*DJH and *VL*JL 454 pyrosequencing was performed on samples from five time points after transmission⁶. Inference of UCA, and identification and production of clone members were performed as described in the Methods (see also Kepler, T. B., manuscript submitted; <http://arxiv.org/abs/1303.0424>). Additional *VH*DJH and *VL*JL genes were identified by 454 pyrosequencing^{6,34,38} and select *VH*DJH and *VL*JL genes were used to produce recombinant antibodies as reported previously³⁴ and described in the Methods. Binding of patient plasma antibodies and CH103 clonal lineage antibody members to autologous and heterologous HIV-1 Env proteins was measured by ELISA and surface plasmon resonance^{19,34,43,45}, and neutralizing activity of patient plasma and CH103 antibody clonal lineage members was determined in a TZM-bl-based pseudovirus neutralization assay^{5,37,46}. Crystallographic analysis of CH103 bound to the HIV-1 outer domain was performed as previously reported⁷, and as described in the Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 2 December 2012; accepted 7 March 2013.

Published online 3 April 2013.

- Walker, L. M. *et al.* Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* **326**, 285–289 (2009).
- Walker, L. M. *et al.* Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477**, 466–470 (2011).
- Burton, D. R., Poignard, P., Stanfield, R. L. & Wilson, I. A. Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science* **337**, 183–186 (2012).
- Kwong, P. D. & Mascola, J. R. Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. *Immunity* **37**, 412–425 (2012).
- Wu, X. *et al.* Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329**, 856–861 (2010).
- Wu, X. *et al.* Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* **333**, 1593–1602 (2011).
- Zhou, T. *et al.* Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* **329**, 811–817 (2010).
- Sattentau, Q. J. & McMichael, A. J. New templates for HIV-1 antibody-based vaccine design. *F1000 Biol. Rep.* **2**, 60 (2010).
- Stamatatos, L. HIV vaccine design: the neutralizing antibody conundrum. *Curr. Opin. Immunol.* **24**, 316–323 (2012).
- Haynes, B. F. *et al.* Cardiolipin polyspecific autoreactivity in two broadly neutralizing HIV-1 antibodies. *Science* **308**, 1906–1908 (2005).
- Haynes, B. F., Kelsoe, G., Harrison, S. C. & Kepler, T. B. B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nature Biotechnol.* **30**, 423–433 (2012).
- Mouquet, H. & Nussenzweig, M. C. Polyreactive antibodies in adaptive immune responses to viruses. *Cell. Mol. Life Sci.* **69**, 1435–1445 (2012).
- Scheid, J. F. *et al.* Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature* **458**, 636–640 (2009).
- Chen, W. *et al.* All known cross-reactive HIV-1 neutralizing antibodies are highly divergent from germline and their elicitation may require prolonged periods of time. *AIDS Res. Human Retrovirol.* **24**, Abstract DA03-03 (2008).
- Dimitrov, D. S. Therapeutic antibodies, vaccines and antibodyomes. *MAbs* **2**, 347–356 (2010).
- Ma, B. J. *et al.* Envelope deglycosylation enhances antigenicity of HIV-1 gp41 epitopes for both broad neutralizing antibodies and their unmutated ancestor antibodies. *PLoS Pathog.* **7**, e1002200 (2011).
- Pancera, M. *et al.* Crystal structure of PG16 and chimeric dissection with somatically related PG9: structure-function analysis of two quaternary-specific antibodies that effectively neutralize HIV-1. *J. Virol.* **84**, 8098–8110 (2010).
- Xiao, X. *et al.* Germline-like predecessors of broadly neutralizing antibodies lack measurable binding to HIV-1 envelope glycoproteins: implications for evasion of immune responses and design of vaccine immunogens. *Biochem. Biophys. Res. Commun.* **390**, 404–409 (2009).
- Alam, S. M. *et al.* Differential reactivity of germ line allelic variants of a broadly neutralizing HIV-1 antibody to a gp41 fusion intermediate conformation. *J. Virol.* **85**, 11725–11731 (2011).
- Bonsignori, M. *et al.* Analysis of a clonal lineage of HIV-1 envelope V2/V3 conformational epitope-specific broadly neutralizing antibodies and their inferred unmutated common ancestors. *J. Virol.* **85**, 9998–10009 (2011).
- Mouquet, H. *et al.* Polyreactivity increases the apparent affinity of anti-HIV antibodies by heterologation. *Nature* **467**, 591–595 (2010).
- Scheid, J. F. *et al.* Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* **333**, 1633–1637 (2011).
- Hoot, S. *et al.* Recombinant HIV envelope proteins fail to engage germline versions of anti-CD4-binding site bNAbs. *PLoS Pathog.* **9**, e1003106 (2013).
- Keele, B. F. *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl Acad. Sci. USA* **105**, 7552–7557 (2008).
- Richman, D. D., Wrinn, T., Little, S. J. & Petropoulos, C. J. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc. Natl Acad. Sci. USA* **100**, 4144–4149 (2003).
- Wei, X. *et al.* Antibody neutralization and escape by HIV-1. *Nature* **422**, 307–312 (2003).
- Corti, D. *et al.* Analysis of memory B cell responses and isolation of novel monoclonal antibodies with neutralizing breadth from HIV-1-infected individuals. *PLoS ONE* **5**, e8805 (2010).
- Gray, E. S. *et al.* The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4⁺ T cell decline and high viral load during acute infection. *J. Virol.* **85**, 4828–4840 (2011).
- Klein, F. *et al.* Broad neutralization by a combination of antibodies recognizing the CD4 binding site and a new conformational epitope on the HIV-1 envelope protein. *J. Exp. Med.* **209**, 1469–1479 (2012).
- Lynch, R. M. *et al.* The development of CD4 binding site antibodies during HIV-1 infection. *J. Virol.* **86**, 7588–7595 (2012).
- Moore, P. L., Gray, E. S. & Morris, L. Specificity of the autologous neutralizing antibody response. *Curr. Opin. HIV AIDS* **4**, 358–363 (2009).
- Moore, P. L. *et al.* Potent and broad neutralization of HIV-1 subtype C by plasma antibodies targeting a quaternary epitope including residues in the V2 loop. *J. Virol.* **85**, 3128–3141 (2011).
- Tomaras, G. D. *et al.* Polyclonal B cell responses to conserved neutralization epitopes in a subset of HIV-1-infected individuals. *J. Virol.* **85**, 11502–11519 (2011).
- Liao, H. X. *et al.* Initial antibodies binding to HIV-1 gp41 in acutely infected subjects are polyreactive and highly mutated. *J. Exp. Med.* **208**, 2237–2249 (2011).
- Tomaras, G. D. *et al.* Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *J. Virol.* **82**, 12449–12463 (2008).
- Scheid, J. F. *et al.* A method for identification of HIV gp140 binding memory B cells in human blood. *J. Immunol. Methods* **343**, 65–67 (2009).
- Seaman, M. S. *et al.* Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for assessment of neutralizing antibodies. *J. Virol.* **84**, 1439–1452 (2010).
- Boyd, S. D. *et al.* Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Sci. Transl. Med.* **1**, 12ra23 (2009).
- Bonsignori, M. *et al.* Two distinct broadly neutralizing antibody specificities of different clonal lineages in a single HIV-1-infected donor: implications for vaccine design. *J. Virol.* **86**, 4688–4692 (2012).
- Haynes, B. F., Moody, M. A., Verkoczy, L., Kelsoe, G. & Alam, S. M. Antibody polyspecificity and neutralization of HIV-1: a hypothesis. *Hum. Antibodies* **14**, 59–67 (2005).
- Wardemann, H. *et al.* Predominant autoantibody production by early human B cell precursors. *Science* **301**, 1374–1377 (2003).
- Shingai, M. *et al.* Most rhesus macaques infected with the CCR5-tropic SHIV(AD8) generate cross-reactive antibodies that neutralize multiple HIV-1 strains. *Proc. Natl Acad. Sci. USA* **109**, 19769–19774 (2012).
- Alam, S. M. *et al.* The role of antibody polyspecificity and lipid reactivity in binding of broadly neutralizing anti-HIV-1 envelope human monoclonal antibodies 2F5 and 4E10 to glycoprotein 41 membrane proximal envelope epitopes. *J. Immunol.* **178**, 4424–4435 (2007).
- Malherbe, D. C. *et al.* Sequential immunization with a subtype B HIV-1 envelope quasiespecies partially mimics the *in vivo* development of neutralizing antibodies. *J. Virol.* **85**, 5262–5274 (2011).
- Alam, S. M. *et al.* Human immunodeficiency virus type 1 gp41 antibodies that mask membrane proximal region epitopes: antibody binding kinetics, induction, and potential for regulation in acute infection. *J. Virol.* **82**, 115–125 (2008).
- Montefiori, D. C. Evaluating neutralizing antibodies against HIV, SIV, and SHIV in luciferase reporter gene assays. *Curr. Protoc. Immunol.* **Chapter 12**, Unit 12.11 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements This study was supported by the National Institutes of Allergy and Infectious Diseases (NIAID) and by intramural National Institutes of Health (NIH) support for the NIAID Vaccine Research Center, by grants from the NIH, NIAID, AI067854 (the Center for HIV/AIDS Vaccine Immunology) and AI100645 (the Center for Vaccine Immunology-Immunogen Discovery). The authors thank J. Pritchett, H. Chen, D. Pause, M. Cooper, E. Solomon, J. Blinn, K. Yarborough, E. Friberg, M. Smith, A. Hogan, C. Peckels, A. Foulger and T. Jeffries for technical assistance, and J. Kircherr

and C. Andrews for project management. Use of sector 22 (Southeast Region Collaborative Access team) at the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract number W-31-109-Eng-38. The opinions herein are those of the authors and should not be construed as official or representing the views of the US Department of Health and Human Services, National Institute for Allergy and Infectious Diseases.

Author Contributions H.-X.L., R.L., T.Z. and F.G. contributed equally to this work. H.-X.L. led production of antibodies and Env proteins, designed assays, analysed data and edited the paper; R.L. generated antibodies and performed assays; T.Z. co-led the structural biology team, performed structural studies, analysed data, and edited the paper; F.G. generated autologous Env sequences and viruses; S.M.A. performed surface plasmon resonance analysis; S.D.B., A.Z.F., J.C.M. and K.M.R. performed pyrosequencing; C.A.S., Z.Z., J.Z. and L.S. analysed pyrosequences; S.G., P.H., B.T. and M.K. performed antibody and Env sequence analysis, and edited the paper; G.K. and G.Y. performed polyreactivity assays and analysis; S.-M.X. and D.C.M. performed neutralization assays and analysis; R.P., K.E.L. and R.M.S. developed and performed ELISAs; K.A.S., M.C. and G.K. performed cohort development, patient recruitment, management and sampling; M.K.L. and L.M.T. performed neutralization assays; Y.C., F.C. and S.C. performed Env cloning and sequencing, S.M., X.D., M.G.J., S.S., B.Z. and A.Z. performed experiments related to crystallization, structure determination, and structural analysis; G.M.S. and B.H.H. generated autologous Env sequences and edited the paper; T.B.K. performed antibody gene sequence analysis and inferred ancestor and intermediate antibodies and edited the paper; P.D.K. co-led the structural biology team and collected and analysed data, and edited the paper; J.R.M. isolated antibodies, designed assays, analysed data, and edited the paper; B.F.H. designed and directed the study, read and interpreted antinuclear antibody assays, analysed data, and wrote and edited the paper.

Author Information The GenBank accession numbers for 292 CH505 Env proteins are KC247375–KC247667, and accessions for 459 V_H DJ_H and 174 V_L J_L sequences of antibody members in the CH103 clonal lineage are KC575845–KC576303 and KC576304–KC576477, respectively. Atomic coordinates and structure factors for unbound CH103 Fab as well as CH103 Fab in complex with the ZM176.66 outer domain have been deposited with the Protein Data Bank under accession codes 4JAM for CH103 Fab, and 4JAN for the CH103–gp120 complex. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.-X.L. (hliao@duke.edu) or B.F.H. (barton.haynes@duke.edu).

NISC Comparative Sequencing Program

Jesse Becker¹, Betty Benjamin¹, Robert Blakesley¹, Gerry Bouffard¹, Shelise Brooks¹, Holly Coleman¹, Mila Dekhtyar¹, Michael Gregory¹, Xiaobin Guan¹, Jyoti Gupta¹, Joel Han¹, April Hargrove¹, Shi-ling Ho¹, Taccara Johnson¹, Richelle Legaspi¹, Sean Lovett¹, Quino Maduro¹, Cathy Masiello¹, Baishali Maskeri¹, Jenny McDowell¹, Casandra Montemayor¹, James Mullikin¹, Morgan Park¹, Nancy Riebow¹, Karen Schandler¹, Brian Schmidt¹, Christina Sison¹, Mal Stantripop¹, James Thomas¹, Pam Thomas¹, Meg Vemulapalli¹ & Alice Young¹

¹NISC Comparative Sequencing Program, NIH, Bethesda, Maryland 20892, USA.

METHODS

Study subject. Plasma and PBMCs were isolated from serial blood samples that were collected from an HIV-1-infected subject CH505 starting 6 weeks after infection up to 236 weeks after infection (Supplementary Table 1) and frozen at -80°C and liquid nitrogen tanks, respectively. During this time, no antiretroviral therapy was administered. All work related to human subjects was in compliance with Institutional Review Board protocols approved by the Duke University Health System Institutional Review Board. Antibodies isolated from PBMCs were tested in binding⁴⁵ and neutralization assays⁴⁶.

Inference of UCA and identification of clone members. The inference of the UCA from a set of clonally related genes is described elsewhere (Kepler, T. B., manuscript submitted; <http://arxiv.org/abs/1303.0424>). In brief, we parameterize the VDJ rearrangement process in terms of its gene segments, recombination points, and *n*-regions sequences (non-templated nucleotides polymerized in the recombination junctions by the action of terminal deoxynucleotidyl transferase). Given any multiple sequence alignment (A) for the set of clonally related genes and any tree (T) describing a purported history, we can compute the likelihood for all parameter values, and thus the posterior probabilities on the rearrangement parameters conditional on A and T. We can then find the unmutated ancestor with the greatest posterior probability, and compute the maximum likelihood alignment A^* and tree T^* given this unmutated ancestor, and then recompute the posterior probabilities on rearrangement parameters conditional on A^* and T^* . We iterate the alternating conditional maximizations until convergence is reached. We use ClustalW⁴⁷ for the multiple sequence alignment, dnaml (PHYLP) to infer the maximum likelihood tree, and our own software for the computation of the likelihood over the rearrangement parameters. The variable regions of heavy- and light-chain (*VH*DJH and *VL*JL) gene segments were inferred from the natural pairs themselves. The posterior probabilities for these two gene segments are 0.999 and 0.993, respectively. We first inferred the unmutated ancestor from the natural pairs as described above. We identified additional clonally related variable region sequences from deep sequencing and refine the estimate of the UCA iteratively. We identified all variable region sequences inferred to have been rearranged to the same *VH*DJH and JH, and to have the correct CDR3 length. For each sequence, we counted the number of mismatches between the sequence and the presumed *VH*DJH gene up to the codon for the second invariant cysteine. Each iteration was based on the CDR3 of the current posterior modal unmutated ancestor. For each candidate sequence, we computed the number of nucleotide mismatches between its CDR3 and the unmutated ancestor CDR3. The sequence was rejected as a potential clone member if the *z*-statistic in a test for difference between proportion is greater than two (ref. 48). Once the set of candidates has been thus filtered by CDR3 distance, the unmutated ancestor was inferred on that larger set of sequences as described above. If the new posterior modal unmutated ancestor differed from the previous one, the process was repeated until convergence was reached. Owing to the inherent uncertainty in unmutated ancestor inference, we inferred the six most likely *VH* UCA sequences resulting in four unique amino acid sequences that were all produced and assayed for reactivity with the transmitted/founder envelope gp140 (Supplementary Table 5).

Phylogenetic trees. Maximum-likelihood phylograms were generated using the dnaml program of the PHYLP package (version 3.69) using the inferred ancestor as the outgroup root, 'speedy/rough' disabled, and default values for the remaining parameters. For the large antibody data sets, neighbour-joining phylogenetic trees were generated using the EBI bioinformatics server (<http://www.ebi.ac.uk/Tools/phylogeny/>) using default parameter values. All neighbour-joining trees were generated subsequent to the inference of the unmutated ancestors.

Isolation and expression of *VH*DJH and *VL*JL genes. The *VH*DJH and *VL*JL gene-segment pairs of the observed CH103, CH104 and CH106 antibodies, and the *VH*DJH gene segment of CH105 were amplified by reverse transcription followed by PCR (RT-PCR) of flow-sorted HIV-1 Env RSC3-specific memory B cells using the methods described previously^{5-7,22,36}. To compare *VH* mutation frequency of CH103, CH104, CH105 and CH106 antibodies with that of previously published of CD4-binding site BnAbs VRC01, CH31 and NIH45-46, *VH* sequences of these antibodies were aligned to the closest *VH* gene segment from the IMGT reference sequence set, and differences between the target sequence and the *VH* gene segment up to and including the second invariant cysteine were counted. The comparison 3' of Cys2 is omitted because the unmutated form of the ancestral sequence is not as well known.

Additional *VH*DJH and *VL*JL genes were identified by 454 pyrosequencing. Clonally related *VH*DJH and *VL*JL sequences derived from either sorted single B cells or 454 pyrosequencing were combined and used to generate neighbour-joining phylogenetic trees (Fig. 2a, b). Antibodies that were recovered from single memory B cells are noted in the figure in red, and bold lines show the inferred evolutionary paths from the UCA to mature BnAbs. For clarity, related *VH*

variants that grouped within monophyletic clades from the same time point were collapsed to single branches, condensing 457 *VH*DJH and 174 *VL*JL variants to 119 and 46 branches, respectively, via the 'nw_condense' function from the Newick Utilities package (v. 1.6)⁴⁹. The frequencies of *VH*DJH variants in each B-cell sample are shown to the right of the *VH*DJH tree in Fig. 2a, and were computed from sample sizes of 188,793, 186,626 and 211,901 sequences from weeks 53, 92 and 144, respectively. Two *VH*DJH genes (IZ95W and 02IV4) were found at 14 weeks after transmission and paired with UCA *VL*JL for expression as IgG1 monoclonal antibodies. The IZ95W monoclonal antibody weakly bound the CH505 transmitted/founder Env gp140 with an end-point titre of $11\ \mu\text{g ml}^{-1}$. Among heavy-chain sequences in the tree, the mean distance of each to its nearest neighbour was calculated to be 8.1 nucleotides. The cumulative distribution function shows that, although there are pairs that are very close together (nearly 30% of sequences are 1 nucleotide from its neighbour), 45% of all sequences differ by 6 nucleotides or more from its nearest neighbour. The probability of generating a sequence that differs by 6 or more nucleotides from the starting sequence by PCR and sequencing is very small. The numbers of sequences obtained from a total of 100 million PBMCs were within the expected range of 50 to 500 antigen-specific B cells.

We have analysed the number of unique *VH*DJH and *VL*JL genes that we have isolated in several ways. First, we have clarified the calculations for the possible number of antigen-specific CD4-binding site memory B cells that could have been isolated from the samples studied. We studied five patient CH505 time points with pyrosequencing, with ~20 million PBMCs per time point for a total of 100 million PBMCs studied. In chronic HIV infection, there is a mean of 145 total B cells per microlitre of blood, and 60 memory B cells per microlitre of blood⁵⁰. This high percentage of memory B cells of ~40% of the total B cells in chronic HIV infection is due to selective loss of naive B cells in HIV infection. Thus, in 100 ml of blood, there will be approximately 6 million memory B cells. If 0.1–1.0% are antigen specific, that would be 6,000–60,000 antigen-specific B cells sampled, and if, of these, 5% were CD4-binding site antibodies, then from 300 to 3000 CD4-binding site B cells would have been sampled in 100 million PBMCs studied. We studied 100 million PBMCs, therefore there should, by these calculations, be 1,000 CD4-binding site B cells sampled. This calculation therefore yields estimates that are completely compatible with the 474 *VH*DJH genes amplified.

To study the plausibility of sequences isolated further, the second method of analysis we used was as follows. Among heavy-chain sequences in the tree, one can compute the distance of each to its nearest neighbour. The mean distance to the nearest neighbour is 8.1 nucleotides. The cumulative distribution function shows that, although there are pairs that are very close together (nearly 30% of sequences are 1 nt from its neighbour), 45% of all sequences differ by 6 nucleotides or more from its nearest neighbour. The probability of generating a sequence that differs by 6 or more nucleotides from the starting sequence by PCR and sequencing is very small. We believe the number of genes represented in our sample is closer to 200 than to 50, and most likely is larger than 200.

The third analysis we performed was to compute the distance of each heavy-chain sequences in the tree to its nearest neighbour. The mean distance to the nearest neighbour is 8.1 nucleotides. We used agglomerative clustering to prune the sequence alignment. At the stage where no pairs of sequences were 3 nucleotides apart or closer, there were 335 out of 452 sequences remaining; when no pairs are 6 nucleotides apart or closer, there are still 288 sequences remaining. Therefore, with this analysis, we believe the number of genes represented in our sample is closer to 300 than to 50, and may be larger. Thus, by the sum of these re-analyses, we believe that the number of genes in the trees in Fig. 2 is plausible.

The isolated Ig *VH*DJH and *VL*JL gene pairs, the inferred UCA and intermediate *VH*DJH and *VL*JL sequences, and select *VH*DJH gene sequences identified by pyrosequencing were studied experimentally (Supplementary Table 2), and used to generate a phylogenetic tree showing the percentage of mutated *VH* sites and time of appearance after transmission (Fig. 2c) and binding affinity (Fig. 2d). The isolated four mature antibodies are indicated in red, antibodies derived from 454 pyrosequencing are indicated in black, and inferred-intermediate antibodies (I1–I4, I7 and I8) are indicated by circles at ancestral nodes. The deep clades in this tree had modest bootstrap support, and the branching order and UCA inference were altered when more sequences were added to the phylogenetic analysis (compare the branching order of Fig. 2a and c). The tree depicted in Fig. 2c, d was used to derive the ancestral intermediates of the representative lineage early in our study, and marked an important step in our analysis of antibody affinity maturation. The *VH*DJH and *VL*JL genes were synthesized (GenScript) and cloned into a pcDNA3.1 plasmid (Invitrogen) for production of purified recombinant IgG1 antibodies as described previously^{51,52}. The *VH*DJH genes of I1–I4, I7 and I8 as well as the *VH*DJH of CH105 were paired with either the *VL* gene of the inferred UCA or I2 depending on the genetic distance of the

VHDJH to either the UCA or mature antibodies for expressing as full-length IgG1 antibodies as described⁵¹ (Supplementary Table 2).

Recombinant HIV-1 proteins. HIV-1 Env genes for subtype B, 63521, subtype C, 1086, and subtype CRF_01_427299, as well as subtype C, CH505 autologous transmitted/founder Env were obtained from acutely infected HIV-1 subjects by single genome amplification²⁴, codon-optimized by using the codon usage of highly expressed human housekeeping genes⁵³, *de novo* synthesized (GeneScript) as gp140 or gp120 (AE.427299) and cloned into a mammalian expression plasmid pcDNA3.1/hygromycin (Invitrogen). Recombinant Env glycoproteins were produced in 293F cells cultured in serum-free medium and transfected with the HIV-1 gp140- or gp120-expressing pcDNA3.1 plasmids, purified from the supernatants of transfected 293F cells by using *Galanthus nivalis* lectin-agarose (Vector Labs) column chromatography^{16,52,54}, and stored at -80°C . Select Env proteins made as CH505 transmitted/founder Env were further purified by superose 6 column chromatography to trimeric forms, and used in binding assays that showed similar results as with the lectin-purified oligomers.

ELISA. Binding of patient plasma antibodies and CH103 clonal lineage antibodies to autologous and heterologous HIV-1 Env proteins was measured by ELISA as described previously^{34,52}. Plasma samples in serial threefold dilutions starting at 1:30 to 1:521,4470 or purified monoclonal antibodies in serial threefold dilutions starting at $100\text{ }\mu\text{g ml}^{-1}$ to $0.000\text{ }\mu\text{g ml}^{-1}$ diluted in PBS were assayed for binding to autologous and heterologous HIV-1 Env proteins. Binding of biotin-labelled CH103 at the subsaturating concentration was assayed for cross-competition by unlabelled HIV-1 antibodies and soluble CD4-Ig in serial fourfold dilutions starting at $10\text{ }\mu\text{g ml}^{-1}$. The half-maximal effective concentration (EC_{50}) of plasma samples and monoclonal antibodies to HIV-1 Env proteins were determined and expressed as either the reciprocal dilution of the plasma samples or concentration of monoclonal antibodies.

Surface plasmon resonance affinity and kinetics measurements. Binding K_d and rate constant (association rate (K_a)) measurements of monoclonal antibodies and all candidate UCAs to the autologous Env C. CH05 gp140 and/or the heterologous Env B.63521 gp120 were carried out on BIAcore 3000 instruments as described previously^{19,43,45}. Anti-human IgG Fc antibody (Sigma Chemicals) was immobilized on a CM5 sensor chip to about 15,000 response units and each antibody was captured to about 50–200 response units on three individual flow cells for replicate analysis, in addition to having one flow cell captured with the control Synagis (anti-RSV) monoclonal antibody on the same sensor chip. Double referencing for each monoclonal antibody–HIV-1 Env binding interactions was used to subtract nonspecific binding and signal drift of the Env proteins to the control surface and blank buffer flow, respectively. Antibody capture level on the sensor surface was optimized for each monoclonal antibody to minimize rebinding and any associated avidity effects. C.CH505 Env gp140 protein was injected at concentrations ranging from 2 to $25\text{ }\mu\text{g ml}^{-1}$, and B.63521 gp120 was injected at $50\text{--}400\text{ }\mu\text{g ml}^{-1}$ for UCAs and early intermediates IA8 and IA4, $10\text{--}100\text{ }\mu\text{g ml}^{-1}$ for intermediate IA3, and $1\text{--}25\text{ }\mu\text{g ml}^{-1}$ for the distal and mature monoclonal antibodies. All curve-fitting analyses were performed using global fit of the 1:1 Langmuir model and are representative of at least three measurements. All data analysis was performed using the BIAevaluation 4.1 analysis software (GE Healthcare).

Neutralization assays. Neutralizing antibody assays in TZM-bl cells were performed as described previously⁵⁵. Neutralizing activity of plasma samples in eight serial threefold dilutions starting at 1:20 dilution and for recombinant monoclonal antibodies in eight serial threefold dilutions starting at $50\text{ }\mu\text{g ml}^{-1}$ were tested against autologous and heterologous HIV-1 Env-pseudotyped viruses in TZM-bl-based neutralization assays using the methods as described^{45,37,55}. Neutralization breadth of CH103 was determined using a previously described^{53,37} panel of 196 of geographically and genetically diverse Env-pseudoviruses representing the major circulated genetic subtypes and circulating recombinant forms. The subtypes shown in Fig. 1c are consistent with previous publications^{5,56}, and the clades described in Los Alamos database (<http://www.hiv.lanl.gov>). HIV-1 subtype robustness is derived from the analysis of HIV-1 clades over time⁵⁷. The data were calculated as a reduction in luminescence units compared with control wells, and reported as IC_{50} in either reciprocal dilution for plasma samples or in micrograms per microlitre for monoclonal antibodies.

Crystallization of antibody CH103 and its gp120 complex. The antigen binding fragment (Fab) of CH103 was generated by LyS-C (Roche) digestion of IgG1 CH103 and purified as previously described⁷. The extended gp120 core of HIV-1 clade C ZM176.66 was used to form a complex with Fab CH103 by using previously described methods⁵⁸. In brief, deglycosylated ZM176.66, constructed as an extended gp120 core⁵⁹, that was produced using the method as described previously⁷ and Fab CH103 were mixed at a 1:1.2 molar ratio at room temperature and purified by size-exclusion chromatography (Hiload 26/60 Superdex S200 prep grade, GE Healthcare) with buffer containing 0.35 M NaCl , 2.5 mM

Tris, pH 7.0 and $0.02\%\text{ Na}_2\text{S}_2\text{O}_3$. Fractions of the Fab or gp120–CH103 complex were concentrated to $\sim 10\text{ mg ml}^{-1}$, flash frozen with liquid nitrogen before storing at -80°C and used for crystallization screening experiments.

Commercially available screens, Hampton Crystal Screen (Hampton Research), Precipitant Synergy Screen (Emerald BioSystems), Wizard Screen (Emerald BioSystems), PACT Suite and JCSG+ (Qiagen) were used for initial crystallization screening of both Fab CH103 and its gp120 complex. Vapour-diffusion sitting drops were set up robotically by mixing $0.2\text{ }\mu\text{l}$ of protein with an equal volume of precipitant solutions (Honeybee 963, DigiLab). The screen plates were stored at 20°C and imaged at scheduled times with RockImager (Formulatrix). The Fab CH103 crystals appeared in a condition from the JCSG+ kit containing 170 mM ammonium sulphate, 15% glycerol and 25.5% PEG 4000. For the gp120–CH103 complex (Supplementary Table 8), crystals were obtained after 21 days of incubation in a fungi-contaminated^{60,61} droplet of the PACT suite that contained 200 mM sodium formate, 20% PEG 3350 and 100 mM bistrispropane, pH 7.5.

X-ray data collection and structure determination for gp120–CH103. Diffraction data were collected under cryogenic conditions. Optimal cryo-protectant conditions were obtained by screening several commonly used cryo-protectants as described previously⁷. X-ray diffraction data were collected at beam-line ID-22 (SER-CAT) at the Advanced Photon Source, Argonne National Laboratory, with $1.0000\text{ }\text{\AA}$ radiation, processed and reduced with HKL2000 (ref. 62). For the Fab CH103 crystal, a data set at $1.65\text{ }\text{\AA}$ resolution was collected with a cryo-solution containing 20% ethylene glycol, 300 mM ammonium sulphate, 15% glycerol and 25% PEG 4000 (Supplementary Table 8). For the gp120–CH103 crystals, a data set at $3.20\text{ }\text{\AA}$ resolution was collected using a cryo-solution containing 30% glycerol, 200 mM sodium formate, 30% PEG 3350 and 100 mM bistrispropane, pH 7.5 (Supplementary Table 8).

The Fab CH103 crystal was in the $P2_1$ space group with cell dimensions at $a = 43.0$, $b = 146.4$, $c = 66.3$, $\alpha = 90.0$, $\beta = 97.7$ and $\gamma = 90.0$, and contained two Fab molecules per asymmetric unit (Supplementary Table 8). The crystal structures of Fab CH103 were solved by molecular replacement using Phaser⁶³ in the CCP4 program suite⁶⁴ with published antibody structures as searching models. The gp120–CH103 crystal also belonged to the $P2_1$ space group with cell dimensions at $a = 48.9$, $b = 208.7$, $c = 69.4$, $\alpha = 90$, $\beta = 107.2$ and $\gamma = 90.0$, and contained two gp120–CH103 complexes per asymmetric unit (Supplementary Table 8). The high-resolution CH103 Fab structure was used as an initial model to place the CH103 Fab component in the complex. With the CH103 Fab position fixed, searching with the extended gp120 core of ZM176.66 in the VRC01-bound form as an initial model failed to place the gp120 component in the complex. After trimming the inner domain and bridging sheet regions from the gp120 search model, Phaser was able to place correctly the remaining outer domain of gp120 into the complex without considerable clashes. Analysis of the packing of the crystallographic lattice indicated a lack of space to accommodate the inner domain of gp120, suggesting possible protease cleavage of gp120 by the containing fungi during crystallization^{60,61}.

Structural refinements were carried out with PHENIX⁶⁵. Starting with torsion-angle simulated annealing with slow cooling, iterative manual model building was carried out on COOT⁶⁶ with maps generated from combinations of standard positional, individual B -factor, TLS (translation/libration/screw) refinement algorithms and non-crystallographic symmetry (NCS) restraints. Ordered solvents were added during each macro cycle. Throughout the refinement processes, a cross validation (R_{free}) test set consisting of 5% of the data was used and hydrogen atoms were included in the refinement model. Structure validations were performed periodically during the model building/refinement process with MolProbity⁶⁷ and pdb-care⁶⁸. X-ray crystallographic data and refinement statistics are summarized in Supplementary Table 8. The Kabat nomenclature⁶⁹ was used for numbering of amino acid residues in amino acid sequences in antibodies.

Protein structure analysis and graphical representations. PISA⁷⁰ was used to perform protein–protein interfaces analysis. CCP4 (ref. 66) was used for structural alignments. All graphical representation with protein crystal structures were made with PyMol⁷¹.

Polyreactivity analysis of antibodies. All antibodies in CH103 clonal lineage were assayed at $50\text{ }\mu\text{g ml}^{-1}$ for autoreactivity to HEP-2 cells (Inverness Medical Professional Diagnostics) by indirect immunofluorescence staining and a panel of autogens by antinuclear antibody assays using the methods as reported previously¹⁰. The intermediate antibody IA1 and CH106 were identified as reactive with HEP-2 cells and then selected for further testing for reactivity with human host cellular antigens using ProtoArray 5 microchip (Invitrogen) according to the instructions of the microchip manufacturer. In brief ProtoArray 5 microchips were blocked and exposed to $2\text{ }\mu\text{g ml}^{-1}$ IA1, CH106 or an isotype-matched (IgG1, k) human myeloma protein, 151 K (Southern Biotech) for 90 min at 4°C . Protein–antibody interactions were detected by $1\text{ }\mu\text{g ml}^{-1}$ Alexa Fluor 647-conjugated anti-human IgG. The arrays were scanned at 635 nm with $10\text{-}\mu\text{m}$ resolution, using

100% power and 600 gain (GenePix 4000B scanner, Molecular Devices). Fluorescence intensities were quantified using GenePix Pro 5.0 (Molecular Devices). Lot-specific protein spot definitions were provided by the microchip manufacturer and aligned to the image.

47. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
48. Zar, J. H. *Biostatistical Analysis* (Pearson, 1974).
49. Junier, T. & Zdobnov, E. M. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* **26**, 1669–1670 (2010).
50. Moir, S. *et al.* Normalization of B cell counts and subpopulations after antiretroviral therapy in chronic HIV disease. *J. Infect. Dis.* **197**, 572–579 (2008).
51. Liao, H. X. *et al.* High-throughput isolation of immunoglobulin genes from single human B cells and expression as monoclonal antibodies. *J. Virol. Methods* **158**, 171–179 (2009).
52. Liao, H. X. *et al.* Vaccine induction of antibodies against a structurally heterogeneous site of immune pressure within HIV-1 envelope protein variable regions 1 and 2. *Immunity* **38**, 176–186 (2013).
53. André, S. *et al.* Increased immune response elicited by DNA vaccination with a synthetic gp120 sequence with optimized codon usage. *J. Virol.* **72**, 1497–1503 (1998).
54. Liao, H. X. *et al.* A group M consensus envelope glycoprotein induces antibodies that neutralize subsets of subtype B and C HIV-1 primary viruses. *Virology* **353**, 268–282 (2006).
55. Montefiori, D. C. *et al.* Magnitude and breadth of the neutralizing antibody response in the RV144 and Vax003 HIV-1 vaccine efficacy trials. *J. Infect. Dis.* **206**, 431–441 (2012).
56. Huang, J. *et al.* Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* **491**, 406–412 (2012).
57. Wolinsky, S. M. *et al.* Response: HIV-1 evolution and disease progression. *Science* **274**, 1010–1011 (1996).
58. Kwong, P. D. *et al.* Probability analysis of variational crystallization and its application to gp120, the exterior envelope glycoprotein of type 1 human immunodeficiency virus (HIV-1). *J. Biol. Chem.* **274**, 4115–4123 (1999).
59. Kwon, Y. D. *et al.* Unliganded HIV-1 gp120 core structures assume the CD4-bound conformation with regulation by quaternary interactions and variable loops. *Proc. Natl Acad. Sci. USA* **109**, 5663–5668 (2012).
60. Bai, Y., Auperin, T. C. & Tong, L. The use of *in situ* proteolysis in the crystallization of murine CstF-77. *Acta Crystallogr. F* **63**, 135–138 (2007).
61. Mandel, C. R., Gebauer, D., Zhang, H. & Tong, L. A serendipitous discovery that *in situ* proteolysis is essential for the crystallization of yeast CPSF-100 (Ydh1p). *Acta Crystallogr. F* **62**, 1041–1045 (2006).
62. Otwinowski, Z. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307 (1997).
63. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
64. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
65. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
66. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
67. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
68. Lütke, T. & von der Lieth, C. W. pdb-care (PDB carbohydrate residue check): a program to support annotation of complex carbohydrate structures in PDB files. *BMC Bioinformatics* **5**, 69 (2004).
69. Kabat, E. A., Wu, T. T., Gottesman, K. S. & Foeller, C. *Sequences of Proteins of Immunological Interest* 5th edn (DIANE publishing, 1991).
70. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
71. DeLano, W. L. The PyMOL Molecular Graphics System <http://www.pymol.org> (DeLano Scientific, 2002).

Accurate assessment of mass, models and resolution by small-angle scattering

Robert P. Rambo¹ & John A. Tainer^{1,2}

Modern small-angle scattering (SAS) experiments with X-rays or neutrons provide a comprehensive, resolution-limited observation of the thermodynamic state. However, methods for evaluating mass and validating SAS-based models and resolution have been inadequate. Here we define the volume of correlation, V_c , a SAS invariant derived from the scattered intensities that is specific to the structural state of the particle, but independent of concentration and the requirements of a compact, folded particle. We show that V_c defines a ratio, Q_R , that determines the molecular mass of proteins or RNA ranging from 10 to 1,000 kilodaltons. Furthermore, we propose a statistically robust method for assessing model-data agreements (χ^2_{free}) akin to cross-validation. Our approach prevents over-fitting of the SAS data and can be used with a newly defined metric, R_{SAS} , for quantitative evaluation of resolution. Together, these metrics (V_c , Q_R , χ^2_{free} and R_{SAS}) provide analytical tools for unbiased and accurate macromolecular structural characterizations in solution.

Achieving reliable, high-throughput structural characterizations of biological macromolecular complexes is a key challenge in the modern structural-genomics era¹. In principle, SAS with X-rays (SAXS) or neutrons (SANS) can meet this challenge by efficiently providing information that fully describes the structural state of a macromolecule in solution^{2–4}. SAS can determine a scattering particle's radius of gyration (R_g), volume (V_p), surface-to-volume ratio and correlation length (l_c), with the latter three physical parameters dependent on the Porod invariant⁵, Q , an empirical SAS value defined for compact folded particles. Q is unique to a scattering experiment and requires convergence of the SAS data at high scattering vectors (q , Å^{−1}) in a $q^2I(q)$ versus q (Kratky) plot. Convergence defines an enclosed area where the degree of convergence reflects the compacted (bounded area), flexible or unfolded (unbounded area) solution states (Fig. 1a). Consequently, non-convergence leaves Q undetermined and paradoxically implies that V_p and l_c are undefined for flexible particles (Supplementary Fig. 1 and Supplementary Notes). This observation leaves R_g as the only structural parameter that can be reliably derived from SAS data on flexible systems.

Defining the V_c

SAS is uniquely capable of providing structural information on all particle types, including flexible systems such as intrinsically unstructured proteins^{6,7}. Here, we overcome current limitations of SAS analyses by deriving a SAS invariant, the V_c . V_c is defined as the ratio of the particle's zero angle scattering intensity, $I(0)$, to its total scattered intensity (Supplementary Notes). The total scattered intensity is the integrated area of the SAS data^{8,9} transformed as $qI(q)$ versus q . Unlike the Kratky plot, we observe that the integral of $qI(q)$ versus q converges for both folded-compact and unfolded-flexible particles (Fig. 1b). The aforementioned ratio at particle concentration (c) and contrast (p) given by

$$V_c = \frac{I(0)}{\int qI(q) dq} = \frac{cV_p^2(\Delta\rho)^2}{cV_p(\Delta\rho)^2 2\pi l_c} = \frac{V_p}{2\pi l_c} \quad (1)$$

reduces to the particle's volume (V_p) per self-correlation length (l_c) with units of Å².

This derivation asserts that V_c , like R_g , can be calculated from a single SAS curve and is concentration independent. We validated concentration independence using well-characterized macromolecules of differing composition and mass. Specifically, for the 173-kDa protein glucose isomerase and the 51-kDa P4–P6 RNA domain from the *Tetrahymena* group I intron¹⁰, SAXS data collected at seven concentrations ranging from 0.2 to 3 mg ml^{−1} showed concentration independence: 86% of the variance was contained within 4% of the mean. Further analysis of seven additional protein and RNA samples confirmed the concentration independence (Fig. 1c): 65% of the variance was contained within 2% of the mean, suggesting that V_c is constant across the concentration ranges for all macromolecular shapes and compositions tested.

V_c is defined by the particle's l_c and implies that a change in conformation should change V_c (Fig. 1d). We observed this prediction for both the bacterial S-adenosylmethionine (SAM)-I riboswitch¹⁰ and Pyr1, a plant hormone-binding protein¹¹. For these macromolecules, ligand binding decreased both R_g and V_c (Table 1), consistent with reported compaction upon binding^{11–13}. Furthermore, we examined Mg²⁺-dependent structured RNAs for folding by SAXS. Measurements of both the bacterial SAM-I riboswitch and turnip yellow mosaic virus (TyMV) transfer-RNA-like structure (TLS)¹⁴ without Mg²⁺ displayed the classic hyperbolic feature of a monodisperse multi-conformation Gaussian ensemble in the Kratky plot (Supplementary Fig. 1). As predicted, flexibility in the absence of Mg²⁺ increased the experimentally determined V_c values (by 14.5% for TyMV TLS and 21% for SAM-I RNA) compared to their compact Mg²⁺-folded states (Table 1). Collectively, the observed ligand-dependent changes in V_c for both Pyr1 and SAM-I RNA or Mg²⁺-dependent changes in V_c for TyMV TLS and SAM-I RNA assert that V_c is an informative descriptor of the macromolecular state.

Particle mass determination by Q_R

Accurate determination of molecular mass has been one of the main difficulties in SAS analysis. Existing methods require an accurate particle concentration, the assumption of a compact near-spherical shape, or SAXS measurements on an absolute scale^{15–18}. As these

¹Life Sciences Division, Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ²Department of Integrative Structural and Computational Biology, The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037, USA.

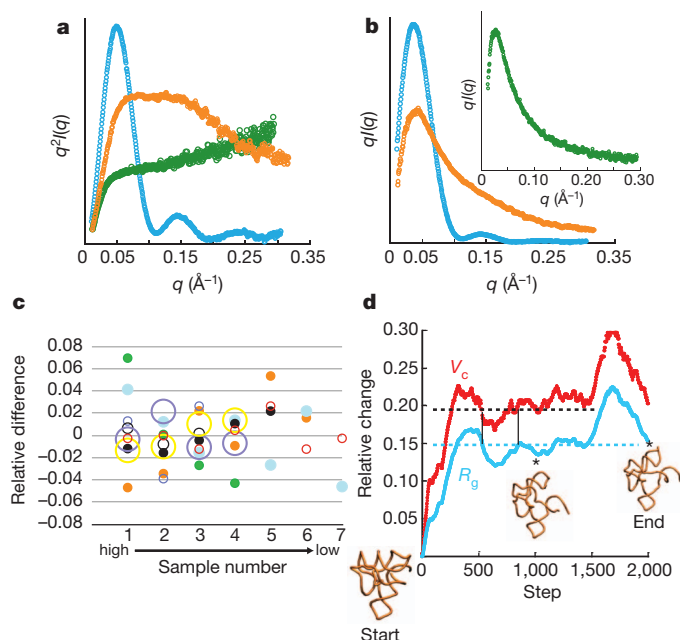


Figure 1 | Concentration independence and conformational dependence of V_c . **a, b**, Experimental SAXS data plotted on a relative scale for glucose isomerase (cyan), 94-nucleotide SAM-I riboswitch in the absence of Mg^{2+} (orange) and human RAD51AP1, an intrinsically unfolded protein (green). **a**, Data transformed as the Kratky plot, $q^2 I(q)$ versus q , reveal the parabolic convergence for a folded particle (blue) and divergence for a flexible (orange) or fully unfolded (green) particle. **b**, Data plotted as $q^2 I(q)$ versus q show convergence for both folded and flexible particles. Inset, convergence for a fully unfolded polymer. **c**, Concentration independence of V_c for experimental SAXS data. For each of nine samples, relative difference is calculated as the deviation from the mean normalized to the mean. Concentrations ranged from 0.2 to 3 mg ml⁻¹ for glucose isomerase (cyan), P4–P6 domain (open red and solid green), xylanase (orange), TyMV TLS RNA (UUAG; solid black), poliovirus del8 competitive inhibitor RNA³¹ (open purple), *Agrobacterium tumefaciens* RNase P (open black), SAM-I riboswitch with Mg^{2+} and ligand (large open purple) and SAM-I riboswitch in the absence of Mg^{2+} (large open yellow). Data was collected to $q_{max} = 0.32 \text{ \AA}^{-1}$ with the exception of solid green ($q_{max} = 0.52 \text{ \AA}^{-1}$). x-axis (sample number) refers to the different concentrations for each sample increasing from left to right. **d**, Correlated changes in V_c (red) and R_g (cyan) for conformations of SAM-I riboswitch (PDB code, 2GIS) simulated from molecular dynamics with CNS²⁸. Horizontal lines demonstrate for R_g or V_c that a single value can map to multiple conformations. Dual specification of both R_g and V_c reduces multiplicity (vertical bars). Relative change represents the difference calculated from the starting model 2GIS. Asterisks denote the time step of the displayed conformation.

Table 1 | Condition-dependent changes in SAXS invariants

Macromolecule	V_c (\AA^3)	R_g (\AA)	V_p^* (\AA^3)	SAXS mass (kDa)
SAM-I (bound): mixture†	460 (± 2)	34.4 (± 0.3)	80,000	50.3
SAM-I (free): mixture	407 (± 2)	31.0 (± 0.2)	76,000	44.9
SAM-I (bound)	280 (± 4)	22.8 (± 0.4)	40,000	31.4
SAM-I (free)	295 (± 4)	24.7 (± 0.7)	48,000	32.0
SAM-I – Mg^{2+}	339 (± 12)	31.6 (± 1.0)	ND	32.8
P4–P6 RNA domain: mixture	478 (± 1.0)	31.0 (± 0.1)	105,000	58.2
P4–P6 RNA domain	414 (± 5)	29.4 (± 0.2)	73,000	50.8
PYR1 (bound)	319 (± 0.5)	20.6 (± 0.9)	59,000	41.9
PYR1 (free)	343 (± 8)	23.2 (± 0.8)	74,000	40.2
TyMV + Mg^{2+}	324 (± 2)	25.9 (± 0.1)	49,000	35.9
TyMV – Mg^{2+}	371 (± 1)	29.9 (± 0.1)	ND	39.8

Uncertainties are the standard deviation of 4–8 independent SAXS data sets.

* V_p denotes the particle's Porod volume.

† Mixture refers to non-gel-filtration-purified samples containing misfolded RNA.

ND, not determined.

requirements hinder both accuracy and throughput of mass estimates by SAS, we sought to establish a SAS-based statistic suitable for determining the molecular mass of proteins, nucleic acids or mixed complexes in solution without concentration or shape assumptions. We calculated R_g and V_c from simulated SAXS profiles for 9,446 protein structures from the Protein Data Bank (PDB)¹⁹, ranging in molecular weight from 8 to 400 kDa. We discovered that a parameter, Q_R , defined as the ratio of the square of V_c to R_g with units of \AA^3 , is linear versus molecular mass in a log–log plot (Figs 2, 3 and Supplementary Fig. 2). The linear relationship is a power-law relationship given by

$$\text{mass} = \left(\frac{Q_R}{e^c} \right)^{1/k} \quad (2)$$

which determines the empirical mass of the scattering biological particle allowing for the direct assessment of oligomeric state and sample quality. Parameters k and c are empirically determined and specific to the class of macromolecular particle (Supplementary Fig. 3), with e as Euler's number.

V_c and R_g are both contrast and concentration independent, thus the determination of molecular mass using Q_R can be made from SAXS data collected under diverse buffer conditions and concentrations, albeit free of inter-particle interference. In fact, this linear relationship produced an average mass error <4% for the 9,446 proteins in the *in vacuo*-simulated data set (Fig. 2).

Calculations of Q_R from simulated and experimental (Supplementary Tables 1 and 2) buffer-subtracted SAXS data of proteins, mixed protein–nucleic acid complexes or RNA alone (Fig. 3a, b) further verified the power-law relationship between Q_R and mass. The mass errors for protein and RNA gel-filtration-purified SAXS samples were

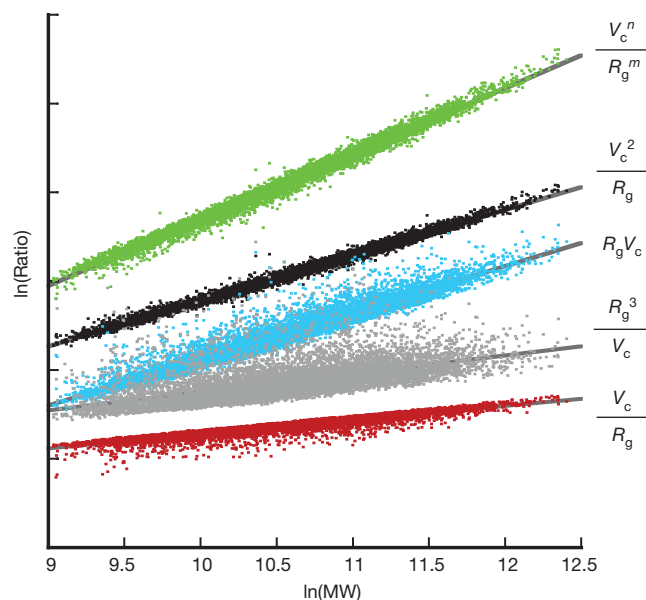


Figure 2 | Defining the power-law relationship between V_c , R_g and protein mass. MW, molecular mass. V_c and R_g were determined from theoretical atomic X-ray scattering profiles for 9,446 protein PDB¹⁹ structures. For each profile, SAXS data were simulated to a maximum $q = 0.5 \text{ \AA}^{-1}$ ($\sim 13 \text{ \AA}$). Various ratios of V_c and R_g against protein mass were examined in a log–log plot. The linear relationship observed for the ratio $V_c^2 R_g^{-1}$ (black) suggests that a power-law relationship exists between the ratio and particle mass of the form $\text{ratio} = c(\text{mass})^k$. The ratio, $V_c^2 R_g^{-1}$, is defined by units of \AA^3 with mass in Daltons. Additional ratios examined (green, cyan, grey and red) displayed asymmetric nonlinear relationships. In green, the fit included generic m (0.9246 ± 0.0008) and n (1.892 ± 0.0005) parameters in a nonlinear surface optimization resulting in an average mass error of $4.9 \pm 4.3\%$. Fitting the linear power-law relationship (black) produces an average mass error of $4.0 \pm 3.6\%$. Truncation of the data to $q = 0.3 \text{ \AA}^{-1}$ ($\sim 21 \text{ \AA}$ resolution) increases the mass error by 0.6% (Supplementary Fig. 2).

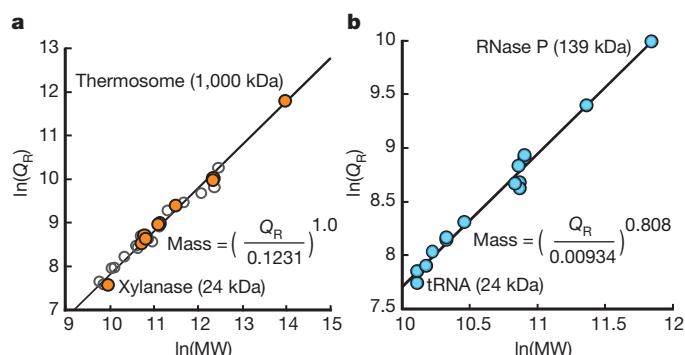


Figure 3 | Power-law relationship between Q_R and particle mass allows direct mass determination. **a**, Q_R calculated from previously reported experimental SAXS data for protein-only samples (Supplementary Table 1). Gel-filtration-purified samples (orange) were plotted with experimental data taken from <http://BioIsis.net> (open circles). **b**, Q_R calculated from experimental SAXS data for RNA-only samples (blue) (Supplementary Table 2). Final equations in **a** and **b** can be used for mass determination of protein- or RNA-only samples. Owing to a lack of available SAXS data for protein-nucleic acid complexes, parameters for k and c remain undetermined.

9.7 and 4.6%, respectively. Furthermore, for RNAs that were measured under folded and unfolded conditions, the average mass difference was 5.6%. The empirically determined mass power-law parameters (Fig. 3) are specific to macromolecular composition and analogous to empirical refractive index increments in light-scattering studies²⁰. Moreover, Q_R , as a mass estimator, assesses SAXS data quality for modelling. For heterogeneous samples, neither R_g nor V_c alone can reliably suggest a corrupted sample. Applying Q_R to P4-P6 and SAM-I RNA samples with known contaminants¹⁰ (Table 1) shows that having 5 and 15% contaminants results in a 14 and 60% mass error, respectively, suggesting that *ab initio* density models would not accurately represent the assumed homogenous solution state.

Cross-validating SAS model-data agreements

Atomistic modelling of SAS data relies on the reduced chi-square (χ^2) error-weighted scoring function^{21,22} that can be unreliable with moderately noisy data sets or over-estimated degrees of freedom (Supplementary Figs 4 and 5). This can lead to over-fitting and model misidentification. In crystallographic and NMR analyses, cross-validation statistical methods mitigate over-fitting and increase confidence in selected model(s)^{23,24}. Here, we present an analogous robust statistical method based on the Nyquist-Shannon sampling and the noisy-channel coding theorems (Supplementary Notes) for evaluating structural models against SAS data.

For a given maximum dimension (d_{\max}), the sampling theorem⁹ determines that the number of unique, evenly distributed observations, n_s , required to represent a particle to a maximum scattering vector (q_{\max}) is given by $(d_{\max} q_{\max})\pi^{-1}$. For example, SAS data to a q_{\max} of 0.3 \AA^{-1} determines for xylanase (d_{\max} , 44 Å) or for 30S ribosomal particle (d_{\max} , 240 Å) that the minimum number of observations is 4 and 23, respectively. This represents a ~20- to 125-fold over-sampling of a SAS curve composed of 500 observations. The Nyquist-Shannon limit (n_s) is the set of maximally independent observations from the band-limited SAS curve (Supplementary Fig. 7). We reasoned that calculating χ^2 from a data set reduced to n_s should more accurately assess the model-data agreement by restricting χ^2 evaluations to the set of independent random variables (Supplementary Notes).

Owing to over-sampling and the uncertainties in q , $I(q)$ and d_{\max} , determining the exact set of Nyquist-Shannon points will be difficult. Nevertheless, application of the noisy-channel coding theorem guarantees noise-free recovery of the SAS signal (Supplementary Notes and Supplementary Figs 8, 9); therefore, we propose the following

sampling procedure for estimating χ^2 that partitions a SAS data set into n_s equal bins for a given d_{\max} . A randomly sampled data point is taken from each bin creating a n_s -length data vector that is used in χ^2 . To minimize outlier influence, χ^2 is taken as the median over k sampling rounds (typically $k = 1,001$) yielding a statistic we call χ^2_{free} . Analogous to R_{free} , χ^2_{free} uses a cross-validation scheme that excludes data from each bin during a round. This technique is akin to the robust least-trimmed squares method²⁵ and provides resistance to outliers, preventing over-fitting and the misidentification of models^{26,27}.

Resisting over-fitting with χ^2_{free}

We tested χ^2_{free} on SAXS data for xylanase at pH 7.2 (Fig. 4a). Based on the fit to the crystallographic structure (PDB code, 1REF; $\chi^2 = 3.9$), SAXS data imply an alternate conformation in solution. Using 1REF as a reference structure, 1,600 conformations were generated and used in a conventional all-data χ^2 determination. Approximately 7% of the

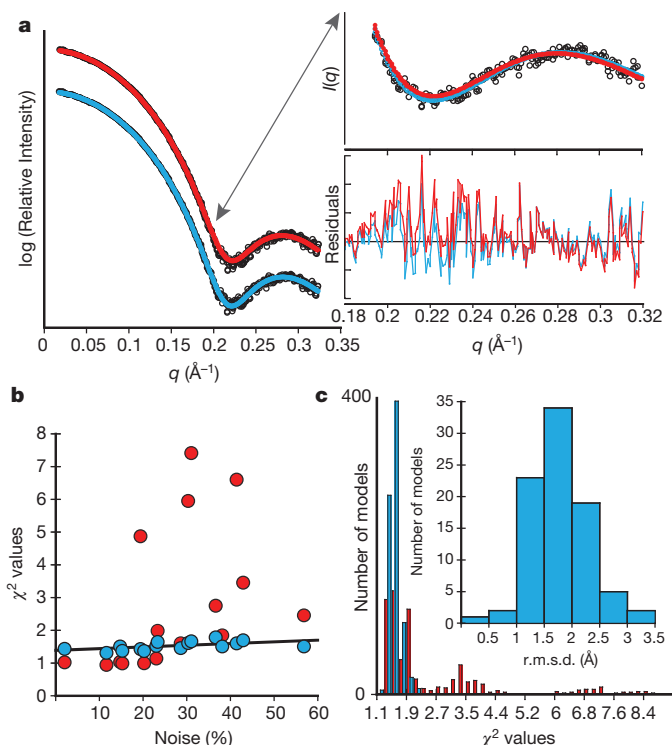


Figure 4 | Objective, quantitative evaluation of models using the least median χ^2 , that is, χ^2_{free} . **a**, Selection of the best PDB model from a pool of 1,600 conformations generated using CONCOORD²⁹. The best selected model (model 44 of 1,600) from CRYSOLO (red) with a conventional $\chi^2 = 1$ demonstrates a bias in the high q region of the residuals, whereas the best selected model (model 560 of 1,600) using χ^2_{free} (cyan) displays an even distribution throughout the residuals with $\chi^2_{\text{free}} = 1.39$. The bias within the high q region ($0.18 \text{ \AA}^{-1} < q < 0.24 \text{ \AA}^{-1}$) implies a conformational difference between the data (red) and target model due to over-fitting. The resistance to over-fitting by χ^2_{free} enables the identification of different 'best' models. **b**, Effects of noise on χ^2 -values from χ^2_{free} (cyan) and conventional χ^2 (red) calculations. Varying empirical noise levels were transposed onto a simulated SAXS profile of a randomly selected xylanase model generated by CONCOORD. A specified noise level represents the average noise in the last third of the q range in **a**. Conventional χ^2 (red) is unstable and directly influenced by outliers producing erroneous χ^2 values, whereas χ^2_{free} is resistant and stable to noise (black line). Erroneous χ^2 values will increase the false-negative rate for an experiment. **c**, Distribution of χ^2 values determined from the set of models with an r.m.s.d. < 1.5 at 19% noise. Thirty randomly selected targets were fitted against 500 simulated SAXS curves at 19% noise from a pool of CONCOORD-generated xylanase conformations. Inset, distribution of r.m.s.d. for all models with a $\chi^2_{\text{free}} < 1.5$. At higher noise, χ^2_{free} (cyan) produces narrower χ^2 -value distributions than conventional χ^2 (red) for near native conformations, thus reducing the overall false-negative rate.

models produced $\chi^2 < 1$, suggesting data over-fitting with the best model ($\chi^2 = 1.0$; Fig. 4a), showing a clear bias in the high q region. Using χ^2_{free} , no model was identified with a $\chi^2_{\text{free}} < 1$ and the best model ($\chi^2_{\text{free}} = 1.39$) demonstrated improved fitting in the high q region, showing that χ^2_{free} distinguishes subtle conformational states. By minimizing on the median n_s -limited χ^2 , χ^2_{free} more accurately determines the true model-data agreement and is not prone to over-fitting (Supplementary Fig. 5).

To test how resistant χ^2_{free} is to noise, we simulated noisy xylanase SAXS data sets using empirical noise from reference data sets and evaluated how well conventional χ^2 and χ^2_{free} can identify the true model from a set of randomly perturbed structures. Under low noise ($\leq 12\%$), both χ^2_{free} and conventional χ^2 behave similarly. At higher noise levels, conventional χ^2 becomes unstable, such that true models would be erroneously rejected. By contrast, χ^2_{free} values were stable over the tested noise levels and effective at identifying matches (Fig. 4b). More importantly, for near-native conformations of the target (root-mean-squared deviation, r.m.s.d. < 1.5), conventional χ^2 values are widely distributed with nearly half greater than 2 (Fig. 4c). For χ^2_{free} , the distribution is narrower, suggesting that near-native conformations are better identified with fewer false negatives.

Validating model-data resolution limits

Determining resolution limits of model-data agreements cannot be achieved by χ^2 alone and requires a metric we define as R_{SAS} , incorporating residuals between modelled and experimental values for both R_g and V_c given by

$$R_{\text{SAS}} = \frac{(R_g^{\text{exp}} - R_g^{\text{model}})^2}{(R_g^{\text{exp}})^2} + \frac{(V_c^{\text{exp}} - V_c^{\text{model}})^2}{(V_c^{\text{exp}})^2} \quad (3)$$

R_{SAS} is a difference distance metric determined from the set of Q -independent SAS invariants. Calculation of R_{SAS} at varying resolutions provides an objective basis to determine appropriate resolution limits for data-model agreements. For dilute xylanase (Supplementary Fig. 4a, b), data were collected to a maximum $q = 0.5 \text{ \AA}^{-1}$ ($\sim 13 \text{ \AA}$ resolution) and fit to PDB 1REF with a χ^2 of 1.3, suggesting an acceptable data-model agreement. However, inspection of R_{SAS} and χ^2_{free} (20.3 and 1.8, respectively) reveal low agreement. Truncating the SAS data shows a significant decrease in R_{SAS} , with χ^2_{free} increasing initially then decreasing as the data-model agreement improves (Supplementary Fig. 4b). Convergence of R_{SAS} towards zero with a $\chi^2_{\text{free}} \leq 1.5$ implies the limit of the data-model agreement to be $q \approx 0.2 \text{ \AA}^{-1}$ or a resolution of 31 \AA . The combination of R_{SAS} and χ^2_{free} for a given model provides a quantitative and graphical approach for determining the acceptable resolution between the data and model (Supplementary Figs 4b and 5). As SAXS data are often used to filter a large set of conformationally distinct models, the models themselves may not be capable of describing the SAXS data to high resolution; therefore, application of R_{SAS} and χ^2_{free} may provide the useful resolution of the data-model agreement. Nevertheless, as done recently for crystallography²⁷, a functional definition of resolution can come from the noisy-channel coding theorem. Here, the useful resolution of the data will be asserted by the highest Nyquist-Shannon point supported by the data.

Perspective

The SAS invariant V_c extends analysis to flexible biopolymers in solution. The volume-per-correlation length, like R_g , faithfully informs on the conformational state of the particle and can be calculated for models determined by other structural techniques including electron microscopy, X-ray crystallography, NMR and SANS. V_c provides a unique descriptor of the scattering experiment that is broadly applicable. We expect that V_c may further characterize voids in materials such as bone, polymeric beads or nanomaterials. As the ratio of the square of V_c to R_g defines a mass parameter, Q_R , SAS experiments can

now inform on particle mass without requiring compactness and instrument calibration. Furthermore, χ^2_{free} is a robust statistical metric that we envision will enable cross-validated determination of flexible ensembles against observed SAXS data. We anticipate that V_c , Q_R , χ^2_{free} and R_{SAS} will efficiently and objectively aid characterization of flexible macromolecules, check sample quality, determine mass and assembly states, detect concentration-dependent scattering, reduce model misidentification and over-fitting, and assess resolution for model-to-data agreement.

METHODS SUMMARY

SAXS data were simulated with FoXS²² and CRY SOL²¹. For each SAXS data set, linear fits to the Guinier region were calculated to determine R_g and $I(0)$. The Guinier parameters were used to calculate an extrapolated scattering data set to zero angle.

On the basis of an extrapolated data set, V_c was calculated by dividing the Guinier $I(0)$ by the integrated area of $qI(q)$ versus q calculated using the trapezoid rule. For simulated atomic SAXS profiles, extrapolation was unnecessary. Simulated atomic SAXS profiles by FoXS calculates scattering profiles at specified scattering-angle increments consistent with experimental measurements, whereas CRY SOL (without an input SAXS data set) can only calculate a maximum of 256 scattering intensities at a specified maximum scattering angle. At beamline 12.3.1 (Advanced Light Source Berkeley), typical data sets collected to a maximum q of 0.32 \AA^{-1} produce ~ 500 data points with the beamstop centred in the middle of the detector. Visual comparison of atomic SAXS profiles from FoXS with CRY SOL did not indicate any systematic differences.

Full Methods and any associated references are available in the online version of the paper.

Received 11 April 2011; accepted 13 March 2013.

- Harrison, S. C. Comments on the NIGMS PSI. *Structure* **15**, 1344–1346 (2007).
- Hura, G. L. *et al.* Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nature Methods* **6**, 606–612 (2009).
- Rambo, R. P. & Tainer, J. A. Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. *Curr. Opin. Struct. Biol.* **20**, 128–137 (2010).
- Sosnick, T. R. & Woodson, S. A. New era of molecular structure and dynamics from solution scattering experiments. *Biopolymers* **95**, 503–504 (2011).
- Glatter, O. & Kratky, O. *Small Angle X-Ray Scattering* (Academic, 1982).
- Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285 (2007).
- Jacques, D. A. & Trewella, J. Small-angle scattering for structural biology—expanding the frontier while avoiding the pitfalls. *Protein Sci.* **19**, 642–657 (2010).
- Bai, Y., Das, R., Millett, I. S., Herschlag, D. & Doniach, S. Probing counterion modulated repulsion and attraction between nucleic acid duplexes in solution. *Proc. Natl Acad. Sci. USA* **102**, 1035–1040 (2005).
- Moore, P. Small-angle scattering. Information content and error analysis. *J. Appl. Crystallogr.* **13**, 168–175 (1980).
- Rambo, R. P. & Tainer, J. A. Improving small-angle X-ray scattering data for structural analyses of the RNA world. *RNA* **16**, 638–646 (2010).
- Nishimura, N. *et al.* Structural mechanism of abscisic acid binding and signaling by dimeric PYR1. *Science* **326**, 1373–1379 (2009).
- Santiago, J. *et al.* Modulation of drought resistance by the abscisic acid receptor PYL5 through inhibition of clade A PP2Cs. *Plant J.* **60**, 575–588 (2009).
- Stoddard, C. D. *et al.* Free state conformational sampling of the SAM-I riboswitch aptamer domain. *Structure* **18**, 787–797 (2010).
- Hammond, J. A., Rambo, R. P. & Kieff, J. S. Multi-domain packing in the aminoacyltransferase 3' end of a plant viral RNA. *J. Mol. Biol.* **399**, 450–463 (2010).
- Orthaber, D., Bergmann, A. & Glatter, O. SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *J. Appl. Crystallogr.* **33**, 218–225 (2000).
- Mylonas, E. & Svergun, D. I. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J. Appl. Crystallogr.* **40**, s245–s249 (2007).
- Fischer, H., de Oliveira Neto, M., Napolitano, H. B., Polikarpov, I. & Craievich, A. F. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Crystallogr.* **43**, 101–109 (2010).
- Rambo, R. P. & Tainer, J. A. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**, 559–571 (2011).
- Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
- Wyatt, P. J. Light scattering and the absolute characterization of macromolecules. *Anal. Chim. Acta* **272**, 1–40 (1993).

21. Svergun, D., Barberato, C. & Koch, M. H. J. CRY SOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
22. Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **38**, W540–W544 (2010).
23. Brünger, A. T. Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475 (1992).
24. Brünger, A. T., Clore, G. M., Gronenborn, A. M., Saffrich, R. & Nilges, M. Assessing the quality of solution nuclear magnetic resonance structures by complete cross-validation. *Science* **261**, 328–331 (1993).
25. Rousseeuw, P. J. & Leroy, A. M. *Robust Regression and Outlier Detection* (Wiley, 1987).
26. Jie, Y., Qi, T., Amores, J. & Sebe, N. *Toward Robust Distance Metric Analysis for Similarity Estimation* (IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006).
27. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
28. Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
29. de Groot, B. L. *et al.* Prediction of protein conformational freedom from distance constraints. *Proteins* **29**, 240–251 (1997).
30. Rambo, R. P. & Doudna, J. A. Assembly of an active group II intron–maturase complex by protein dimerization. *Biochemistry* **43**, 6486–6497 (2004).
31. Keel, A. Y., Jha, B. K. & Kieft, J. S. Structural architecture of an RNA that competitively inhibits RNase L. *RNA* **18**, 88–89 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank G. L. Hura, M. Hammel, R. T. Batey, J. Tanamachi and the staff of SIBYLS Beamline 12.3.1 at the Advanced Light Source for discussions and P. Adams for suggestions regarding simulations with CNS. We thank E. Rambo, G. Williams and E. D. Getzoff for manuscript comments. This work is supported in part by funding to foster collaboration with Bruker and the Berkeley Laboratory Directed Research and Development (LDRD) program provided by the Director, Office of Science, US Department of Energy on Novel Technology for Structural Biology. The SIBYLS Beamline 12.3.1 facility and team at the Advanced Light Source is supported by United States Department of Energy program Integrated Diffraction Analysis Technologies (DEAC02-05CH11231) and by National Institutes of Health grant R01GM105404.

Author Contributions R.P.R. developed the theory and computational algorithms with input from J.A.T. Both J.A.T. and R.P.R. designed the experiments and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.P.R. (rprambo@lbl.gov) or J.A.T. (jat@scripps.edu).

METHODS

χ^2 free calculation. For a given d_{\max} , the SAXS/SANS data collected between q_{\min} and q_{\max} can be divided into n_s equal bins, in which n_s is determined by the Nyquist–Shannon sampling theorem⁹. Here, d_{\max} is measured from the atomistic model; however, d_{\max} can be directly inferred using an indirect Fourier transform method such as GNOM. In the case of 500 data points, and $n_s = 10$, each bin will contain 50 data points such that a single randomly selected data point will represent that Nyquist–Shannon point. As a selected data point may be biased by inter-particle interference or uncertainties in q or $I(q)$, the selection of the representative data point from the Nyquist–Shannon bin must occur through several selection rounds (k). During each round, the set of randomly selected points comprises the test set for calculating χ^2 against the model. The accepted value is taken as the median over k rounds. The number of rounds, k , will vary with the average noise level of the SAXS/SANS data set. The probability of selecting an erroneous data point from a bin scales directly with the noise. We have found that for high-quality data (<10% noise) k can be as small as a few hundred, whereas for high-noise data, k should be 2,000 to a maximum of 3,000.

Sample preparation. Protein and RNA samples were derived from a variety of sources. For glucose isomerase and xylanase, protein samples were obtained as suspended crystals (Hampton Research). Each protein was further purified by gel-filtration chromatography immediately before SAXS data collection in buffer containing either buffer A (20 mM HEPES, pH 7.2, 5 mM MgCl₂, 100 mM KCl and 2 mM tris(2-carboxyethyl)phosphine (TCEP)), buffer B (40 mM 2-ethanesulfonic acid (MES), pH 6.8, 8 mM MgCl₂ and 100 mM KCl) or buffer C (40 mM Na-citrate, pH 5.0, 75 mM KCl and 1% glycerol). Proteins were re-suspended by a 50-fold dilution of the crystals in buffers A or B for glucose isomerase and buffers A, B or C for xylanase. Diluted crystals were incubated at 37 °C on a nutator for 1 h, concentrated to 10 mg ml⁻¹ and injected on a pre-equilibrated Superdex 200 PC 3.2 column (GE Healthcare) for glucose isomerase and Superdex 75 PC 3.2 column (GE Healthcare) for xylanase. Fractions corresponding to peak elution were taken for SAXS and quantitated by absorbance at 280 nm.

Taq polymerase was recombinantly expressed and purified from *Escherichia coli* using cells transformed with a pET vector conferring ampicillin resistance. Cells were grown at 37 °C and induced for 4 h with isopropyl- β -D-thiogalactoside at $D_{260\text{ nm}} = 0.8$ before collection. Cells were lysed as described³⁰. Lysate was clarified by low-speed spin in 50-ml Falcon tubes and incubated at 65 °C for 20 min. Lysate was further clarified by high-speed centrifugation at 20,000g for 40 min at 4 °C. Bound nucleic acids were removed by polyethyleneimine treatment and ammonium sulphate precipitation. Protein was re-suspended in buffer B and further purified to homogeneity using Superdex 200 HR 10/30 (GE Healthcare) for SAXS analysis.

Catalase (human erythrocyte) was purchased from a commercial source (EMD). One milligram was re-suspended in 100 μ l of buffer A and further purified using a Superose 6 PC 3.2 column (GE Healthcare) equilibrated in buffer A. Fraction corresponding to peak elution was taken for SAXS analysis.

Thermosome from *Sulfolobus solfataricus* was purified from source and provided by S. Yannoni. Thermosome samples were prepared by purification on a Superose 6 HR 10/30 column in buffer equilibrated with 40 mM, pH 5.5, 75 mM KCl, 75 mM NaCl, 5 mM MgCl₂, and 2 mM TCEP. Fraction corresponding to peak elution was taken for SAXS analysis.

Data for full-length and truncated Tbl1 was provided by Y. Dimitrova and W. Chazin. Data for p65 was kindly provided by A. Berman and T. Cech. Data for Pyr1 samples were provided by K. Hitomi and E. Getzoff and purified as described¹¹. Samples were purified and analysed on-site by gel-filtration and multi-angle light scattering (MALS) immediately before SAXS analysis.

MALS. MALS studies were performed in line with size-exclusion chromatography on protein and RNA samples to assess monodispersity and mass of the SAXS samples using an 18-angle DAWN HELEOS light-scattering detector in which detector 12 was replaced with a DynaPro quasi-elastic light-scattering detector (Wyatt Technology). Simultaneous concentration measurements were made with an Optilab rEX refractive index detector (Wyatt Technology) connected in tandem to the light-scattering detector. For each buffer used, the MALS system was calibrated with BSA at 10 mg ml⁻¹ to determine delay times and band broadening. For proteins, BSA, xylanase and glucose isomerase provided an additional calibration of the refractive index increment for protein samples. For RNA samples, the refractive index increment was determined from P4–P6 RNA samples^{10,30}.

MALS analyses were performed on all the RNAs (except tRNA^{Phe}) in this study and a set of proteins comprising glucose isomerase, xylanase, thermosome, catalase, Tbl1, Pyr1 and p65 (Supplementary Tables 1 and 2).

PDB query. The PDB was used as a source for structural models for SAXS simulations. The comprehensive protein data set was selected on the basis of the following criteria: molecular mass range, 10–1,200 kDa; technique, X-ray crystallography; resolution limits, 1.8–3.2 Å; exclude 90% similarity; protein only;

single models with one to two chains in the asymmetric unit. Further manual curation was performed for structures in which the asymmetric unit produced two models physically separated in space without crystal contacts. For the RNA-only data sets, the following criteria was used: RNA only; molecular mass range, 10–250 kDa; exclude 95% similarity; technique, X-ray crystallography; single model. Finally, for mixed protein–nucleic acid complexes, the following criteria was used: molecular mass range, 8–1,000 kDa; technique, X-ray crystallography; protein and RNA; protein and DNA; 95% similarity; single model.

SAXS data collection. SAXS data were collected at beamline 12.3.1 of the Advanced Light Source at the Lawrence Berkeley National Laboratory². SAXS data were collected as a two-thirds dilution series using 20- μ l samples and three different exposures. Exposures generally follow a short, medium and long time consisting of 0.1, 1 and 6 s or 0.5, 1 and 8 s, and were merged as described¹⁰. Samples after gel-filtration purification eluted within the range of 1.5 and 3 mg ml⁻¹, and for each sample buffer was collected from the gel-filtration column after 1.2 column volumes for corresponding matching SAXS buffers.

For each sample, aggregation and inter-particle interference was assessed using overlay plots of the concentration series in GnuPlot (<http://www.gnuplot.org>). Fits to the Guinier region ($qR_g < 1.3$) were performed with software at beamline 12.3.1 (R.P.R.) and all data graphs were prepared with Kaleidagraph (<http://www.synergy.com>) and GnuPlot. Figures with structural models were prepared with VMD and rendered with Povray (<http://www.povray.org>).

SAXS data analysis. For each SAXS data set used in this study, linear fits to the Guinier region were performed with rubyGSL (by Y. Tsunesada) and the GNU Scientific Library (<http://www.gnu.org/software/gsl/>) for the determination of R_g and $I(0)$. The Guinier parameters were subsequently used to calculate an extrapolated scattering data set to zero angle at intervals determined from the average scattering vector increment, Δq .

Based on an extrapolated data set, V_c was calculated by dividing the Guinier $I(0)$ by the area of the transformed intensity taken as the product of $qI(q)$ and integrating using the trapezoid rule. For simulated atomic SAXS profiles, extrapolation was not necessary. Simulated atomic SAXS profiles were calculated with FoXS as it can calculate scattering profiles at specified scattering vector increments consistent with experimental measurements whereas CRY SOL (without an input SAXS data set) can only calculate a maximum of 256 scattering intensities at a specified maximum scattering vector. Typical data sets collected at a maximum q of 0.32 Å⁻¹ at beamline 12.3.1 produce ~500 data points with the beamstop centred in the middle of the detector. Visual comparison of atomic SAXS profiles from FoXS with CRY SOL did not illustrate any systematic differences.

For experimental SAXS data sets that were fit to an input PDB model, CRY SOL was used with default input parameters. In these cases, CRY SOL reports χ and not χ^2 for the model fits in the output log file.

Conformational simulation. SAM-I riboswitch molecular dynamics simulations were performed with CNS as described¹³. In brief, the SAM crystal structure (PDB code, 2GIS) was analysed with FIRST and FRODA³² at several energy cutoffs to determine plausible rigid and flexible regions within the structure. These were used to ascribe constraints within the structure for molecular dynamic simulations with CNS using anneal.inp. The CNS input file was modified to remove the electrical potential from the energy function and calculations were performed as torsional angle dynamics only. For each simulation, 2,000 steps were recorded in the trajectory file and each step was written to file as a PDB.

CONCOORD simulations with IREF were performed with the command line argument 'disco -op disco -n 1000 -bump -damp 2 -viol 5 -t 100' to generate 1,000 possible conformations close to the starting input structure. The resulting PDB files were fit to the experimental SAXS data set with CRY SOL and the output intensity file for each PDB conformation was used to calculate V_c .

Simulating noisy SAXS data sets. SAS intensities over a single exposure will range over several decades, and consequently the noise levels will vary throughout the measured q region. Therefore, we used intensity uncertainties from previously collected SAXS experiments as a source of realistic noise for the simulated SAXS data sets. The noise level of the empirical SAXS curve is reported as the average relative noise in the last third of the observed q -range (Fig. 4).

For a selected q , the simulated $I(q)$ was randomly displaced based on a random draw using the Box–Muller transform of a standard Gaussian distribution parameterized by the empirical intensity, $I(q)_{\text{obs}}$, and uncertainty, $\text{error}(q)_{\text{obs}}$. The Box–Muller transform returns two possible values and a random binary selection was used to provide a final single value for the displacement of the simulated $I(q)$, $I(q)_{\text{displaced}}$. The simulated $\text{error}(q)$ was reported as $I(q)_{\text{displaced}} * \text{error}(q)_{\text{obs}}/I(q)_{\text{obs}}$.

32. Fulle, S. & Gohlke, H. Analyzing the flexibility of RNA structures by constraint counting. *Biophys. J.* **94**, 4202–4219 (2008).

Experimental realization of non-Abelian non-adiabatic geometric gates

A. A. Abdumalikov Jr¹, J. M. Fink^{1†}, K. Juliusson^{1†}, M. Pechal¹, S. Berger¹, A. Wallraff¹ & S. Filipp¹

The geometric aspects of quantum mechanics are emphasized most prominently by the concept of geometric phases, which are acquired whenever a quantum system evolves along a path in Hilbert space, that is, the space of quantum states of the system. The geometric phase is determined only by the shape of this path^{1–3} and is, in its simplest form, a real number. However, if the system has degenerate energy levels, then matrix-valued geometric state transformations, known as non-Abelian holonomies—the effect of which depends on the order of two consecutive paths—can be obtained⁴. They are important, for example, for the creation of synthetic gauge fields in cold atomic gases⁵ or the description of non-Abelian anyon statistics^{6,7}. Moreover, there are proposals^{8,9} to exploit non-Abelian holonomic gates for the purposes of noise-resilient quantum computation. In contrast to Abelian geometric operations¹⁰, non-Abelian ones have been observed only in nuclear quadrupole resonance experiments with a large number of spins, and without full characterization of the geometric process and its non-commutative nature^{11,12}. Here we realize non-Abelian non-adiabatic holonomic quantum operations^{13,14} on a single, superconducting, artificial three-level atom¹⁵ by applying a well-controlled, two-tone microwave drive. Using quantum process tomography, we determine fidelities of the resulting non-commuting gates that exceed 95 per cent. We show that two different quantum gates, originating from two distinct paths in Hilbert space, yield non-equivalent transformations when applied in different orders. This provides evidence for the non-Abelian character of the implemented holonomic quantum operations. In combination with a non-trivial two-quantum-bit gate, our method suggests a way to universal holonomic quantum computing.

A cyclic evolution of a non-degenerate quantum system is in general accompanied by a phase change of its state. The acquired Abelian phase can be divided into two parts: the dynamical phase, which is proportional to the evolution time and the energy of the system, and the geometric phase, which depends only on the path of the system in Hilbert space. If the system is guided along two different loops in a row, the overall accumulated geometric phase is independent of their sequential arrangement, because the Abelian phases associated with the loops are additive. The situation changes drastically if the energy spectrum of the system contains degenerate subspaces. In this case, the system can undergo a path-dependent unitary transformation by acquiring a non-Abelian holonomy. Because the holonomic transformations do not commute, the order of the two successive loops makes a difference.

The ability to realize non-commuting quantum operations by choosing different paths can be employed for holonomic quantum computation⁸, which has attracted particular attention¹⁶ because of the resilience of geometric phases to certain fluctuations during the evolution of the system^{17–20}. In this scheme, quantum bits are encoded in a doubly degenerate eigenspace of the system Hamiltonian, $h(\lambda)$. The parameters encoded in the vector λ are varied to induce a cyclic evolution of the system. In the original proposal on holonomic

quantum computation⁸, the parameters λ are changed adiabatically in time to guarantee the persistence of the degeneracy. Adiabatic holonomic gates have been proposed for trapped ions⁹, superconducting quantum bits^{21,22} (qubits) and semiconductor quantum dots²³. However, they are difficult to realize experimentally because of the long evolution time needed to fulfil the adiabatic condition. Instead, a scheme based on non-adiabatic, non-Abelian holonomies¹³ has been proposed¹⁴. Such holonomies, like non-adiabatic geometric gates²⁴, combine universality and speed and can thus be implemented conveniently in experiments.

The main idea is to generate a non-adiabatic and cyclic state evolution in a three-level system that results in a purely geometric operation on the degenerate subspace spanned by the computational basis states, $|0\rangle$ and $|1\rangle$. The third state, $|e\rangle$, acts as an auxiliary state and remains unpopulated after the gate operation. This is achieved by driving the system using two resonant microwave pulses (Fig. 1a) with identical time-dependent envelopes, $\Omega(t)$, but different amplitudes, a and b , satisfying $|a|^2 + |b|^2 = 1$ (Fig. 1b). The Hamiltonian of the system in the interaction picture is

$$h(t) = \frac{\hbar\Omega(t)}{2} (a|e\rangle\langle 0| + b|e\rangle\langle 1| + \text{H.c.})$$

where \hbar denotes Planck's constant divided by 2π , H.c. stands for Hermitian conjugate and we have used the rotating-wave approximation. This Hamiltonian causes the initial basis vectors to evolve to the states $|\psi_i(t)\rangle = \exp\left(-\frac{i}{\hbar} \int_0^t h(t') dt'\right) |i\rangle$ ($i, j = 0, 1$). Unlike in adiabatic schemes, the $|\psi_i(t)\rangle$ are not instantaneous eigenstates of $h(t)$. By keeping a/b , the complex amplitude ratio of the pulses, constant, no transitions between states are induced and the evolution satisfies the parallel-transport condition, $\langle\psi_i(t)|h(t)|\psi_j(t)\rangle = 0$. As a consequence, the evolution is purely geometric, with vanishing dynamic contributions (Supplementary Information). If the pulse length, τ , is chosen such that $\int_0^\tau \Omega(t) dt = 2\pi$, the degenerate subspace undergoes a cyclic evolution, and the matrix representation of the final operator that acts on the basis states $|0\rangle$ and $|1\rangle$ is

$$U = \begin{pmatrix} \cos(\theta) & e^{i\phi} \sin(\theta) \\ e^{-i\phi} \sin(\theta) & -\cos(\theta) \end{pmatrix} \quad (1)$$

where we have parameterized the drive amplitudes via the relation $e^{i\phi} \tan(\theta/2) = a/b$. A geometric interpretation of the dynamics of the system is visualized in Fig. 1c, in which different values of θ and ϕ correspond to different paths \mathcal{C} in Hilbert space.

In our experiments, we realize the holonomic gates using a three-level superconducting artificial atom of the transmon type, embedded in a three-dimensional cavity²⁵ (Fig. 2a). The cavity is made of aluminium and has inner dimensions of $32 \text{ mm} \times 15.5 \text{ mm} \times 5 \text{ mm}$. The frequency of the fundamental mode is $\omega_{\text{ro}}/2\pi \approx 8.999 \text{ GHz}$ as measured by transmission spectroscopy using the circuit shown in Fig. 2b. The quality factor of the resonator is $Q = 21,000$. The transmon is made of two $500 \mu\text{m} \times 250 \mu\text{m}$ aluminium electrodes separated by

¹Department of Physics, ETH Zürich, CH-8093 Zürich, Switzerland. [†]Present addresses: Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, California 91125, USA (J.M.F.); Quantronics Group, Service de Physique de l'Etat Condensé (CNRS, URA 2464), IRAMIS, CEA-Saclay, 91191 Gif-sur-Yvette, France (K.J.).

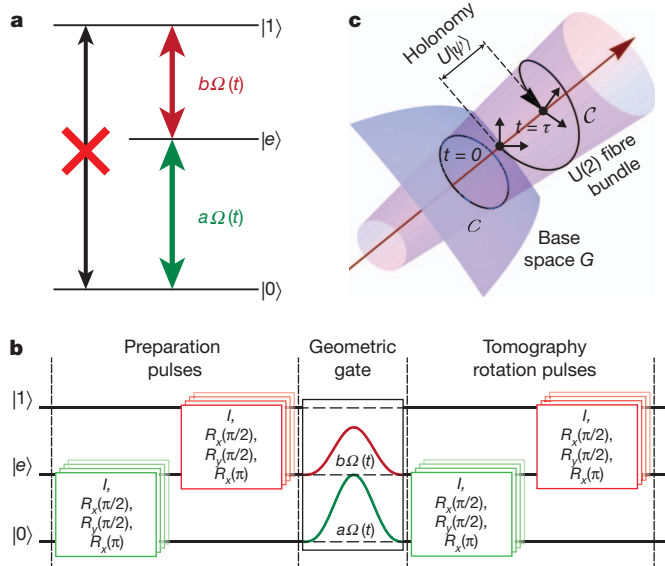


Figure 1 | Geometric gate operation on a three-level system. **a**, Two drive tones with amplitudes a and b and pulse envelope $\Omega(t)$ couple the states $|0\rangle$ and $|1\rangle$ to the state $|e\rangle$. To lowest order, the direct transition between $|0\rangle$ and $|1\rangle$ is forbidden. **b**, Pulse sequence for process tomography. The input state is prepared by sequentially applying pulses on the $|0\rangle \leftrightarrow |e\rangle$ and $|e\rangle \leftrightarrow |1\rangle$ transitions (Methods). The holonomic gate is formed by the simultaneous application of two pulses with envelope $\Omega(t)$ and amplitudes a and b . A set of pulses on the $|0\rangle \leftrightarrow |e\rangle$ and $|e\rangle \leftrightarrow |1\rangle$ transitions followed by the measurement completes the process tomography. **c**, Holonomic gate represented on a fibre bundle. The system evolves along a closed path, C , in the base space, G , of a fibre bundle¹⁴. The base space is the set of two-dimensional degenerate subspaces within the Hilbert space spanned by the state vectors $\{|0\rangle, |1\rangle, |e\rangle\}$. The subspaces remain unchanged under a unitary basis transformation; equivalent subspaces are represented by a fibre associated to each point in G . The parallel-transport condition fixes the choice of basis states along the loop C , which leads to the path C in the fibre bundle. The difference between the initial ($t = 0$) and final ($t = \tau$) points lying on a single fibre corresponds to the holonomy matrix $U \in U(2)$, which is fully determined by the loop C .

130 μm and connected by a Josephson junction (Fig. 2c, d). It is oriented parallel to the electric field of the fundamental mode pointing along the smallest dimension of the cavity (Fig. 2a). The measured coupling strength between the transmon and the cavity field is $g/2\pi \approx 110$ MHz. To read out the state of the transmon, we measure the state-dependent transmission through the cavity²⁶. The transition frequencies measured by Ramsey spectroscopy are $\omega_{0e}/2\pi \approx 8.086$ GHz ($|0\rangle \leftrightarrow |e\rangle$) and $\omega_{e1}/2\pi \approx 7.776$ GHz ($|e\rangle \leftrightarrow |1\rangle$). The decay time of both excited states is $T_1 = 7 \pm 0.1$ μs , and the respective dephasing times are $T_2^{0e} = 8.0 \pm 0.1$ μs and $T_2^{e1} = 3.9 \pm 0.1$ μs .

Different holonomic gates are realized by adjusting the ratio of the two drives, $a/b = e^{i\phi} \tan(\theta/2)$, to values between 0 and $-1/\sqrt{2}$, with $\phi = \pi$. In geometric terms, the change of the transformation from the phase-flip gate, denoted by the Pauli operator σ_z , to the NOT gate, σ_x , is caused by the deformation of the loop C (Fig. 1c). The envelopes $\Omega(t)$ are truncated Gaussian pulses²⁷ with a width of $\sigma = 10$ ns and a total pulse length of 40 ns. With pulses of this length, off-resonant driving of neighbouring transitions is small (Supplementary Information). The performance of the holonomic gates is characterized by process tomography using all three basis states (Methods). The diagonal elements of the reduced process matrix, $\tilde{\chi}$, are shown in Fig. 3a as a function of θ , with $\phi = \pi$. The experimentally obtained results are in good agreement with theory. For $\theta = 0$, a single drive on the $|e\rangle \leftrightarrow |1\rangle$ transition causes a phase shift corresponding to $\tilde{\chi}_{ZZ} = 1$ and the operation $U(\theta = 0) = \sigma_z$. For $\theta = \pi/4$, the transformation $H = (\sigma_z - \sigma_x)/\sqrt{2}$ is generated (Fig. 3b), which is equivalent to the Hadamard gate. For $\theta = \pi/2$, the pulses are of equal amplitude and the transformation is a NOT gate, σ_x (Fig. 3c). Because of dephasing and relaxation of both excited states as

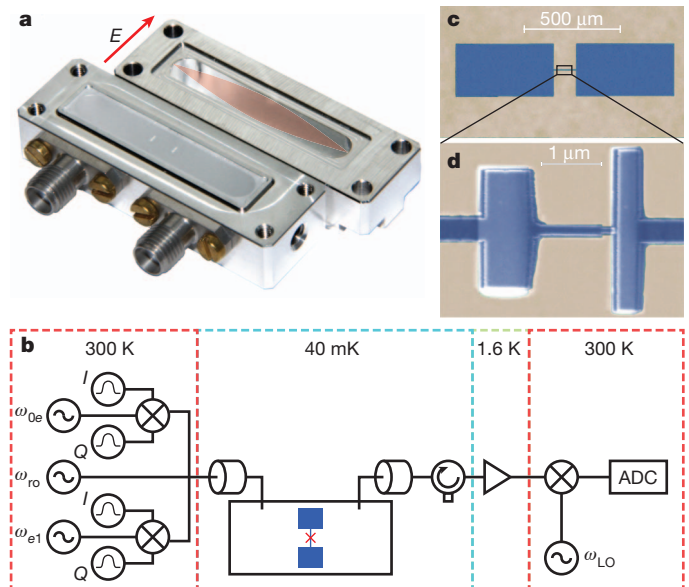


Figure 2 | Transmon qubit in a cavity resonator. **a**, Aluminium cavity with an embedded sapphire chip containing two transmon qubits. The left qubit is not used in the experiment. The electric field profile, E , of the fundamental mode is sketched (red) in the upper part of the cavity. **b**, Microwave control pulses at the transition frequencies ω_{0e} and ω_{e1} are created by modulating two continuous signals using in-phase/quadrature (I/Q) mixers. Modulation pulses in the I and Q channels are generated using arbitrary waveform generators. The control pulses and the read-out pulse at frequency ω_{ro} are combined and fed into the cavity. The transmitted signal is amplified and detected in a heterodyne measurement at room temperature with a local oscillator signal at frequency ω_{LO} , and analysed on a computer using a digitizer (ADC). **c**, Optical micrograph of the transmon device that we used. **d**, Scanning electron micrograph of the Josephson junction.

well as the finite fidelities of the microwave pulses, the state $|e\rangle$ is slightly populated after the gate operation. This population ‘leakage’ is quantified by computing the trace $\text{Tr}(\tilde{\chi}) \approx 0.96$ of the reduced process matrix $\tilde{\chi}$ (Fig. 3a, black dots).

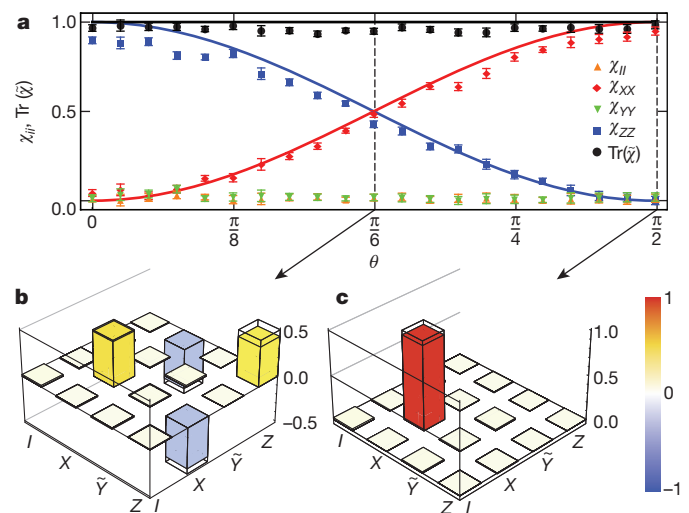


Figure 3 | Process tomography of holonomic gates. **a**, Diagonal elements, χ_{ii} , of the process matrices for ideal (lines) and experimental (symbols) geometric gates as a function of θ , with $\phi = \pi$. Dots correspond to the trace of the reduced process matrix, $\tilde{\chi}$. Error bars are estimated from the Gaussian distribution inferred from the tomography of the final states before performing the maximum-likelihood procedure. **b**, **c**, Bar charts of the real parts of the measured reduced process matrices of the geometric Hadamard gate H (**b**) and the NOT gate (**c**). The wire frames show the theoretically expected values. The small ($\sim 0.3\%$) imaginary parts of $\tilde{\chi}$ are not shown.

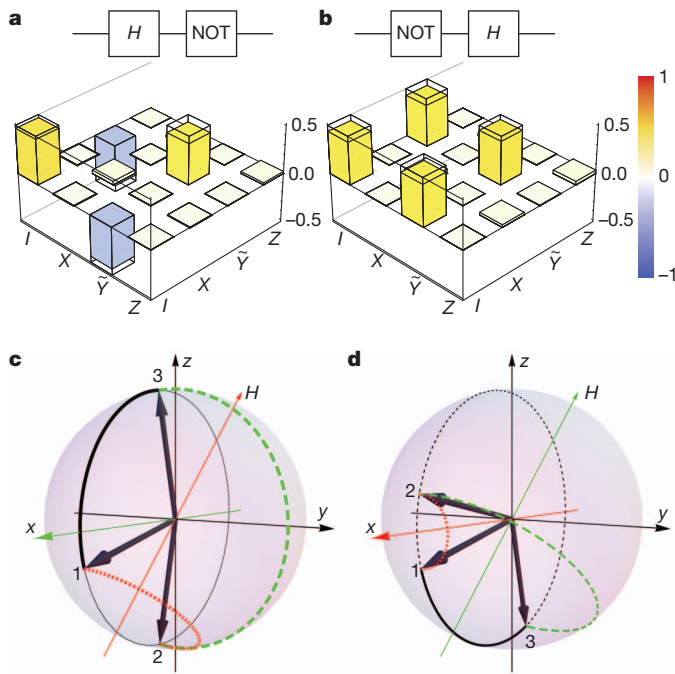


Figure 4 | Non-commutativity of holonomic gates. **a, b**, Process matrices for NOT · H (**a**) and H · NOT (**b**) gates with fidelities of 95%. Because of the non-Abelian character of the geometric operations, the resulting processes are different. **c, d**, This can be visualized on the Bloch sphere by two rotations around the x and H axes (**a** represented by **c**; **b** represented by **d**). The initial state of the system (1) is rotated around the red axis first (red dotted line) towards the intermediate state (2), and is then rotated around the green axis (green dashed line) to the final state (3). The effect of the combined rotations is shown by the thick black line.

The experimentally obtained fidelities of the geometric transformations are $F_H = 95.4 \pm 0.6\%$ and $F_{\text{NOT}} = 97.5 \pm 0.9\%$. The numerical solution of a master equation including dissipative processes results in a fidelity of $F = 97.6\%$ for both processes, in good agreement with the experimental values. From this, we conclude that decoherence and decay processes along with dynamical contributions (Supplementary Information) are the main limiting factors for gate performance.

To show explicitly that different loops in Hilbert space result in non-commuting quantum gates, we sequentially apply the geometric transformations H and NOT in different orders. The non-Abelian character of the operation yields either the operation NOT · H = $-(i\sigma_y + I)/\sqrt{2}$ (Fig. 4a) or H · NOT = $(i\sigma_y - I)/\sqrt{2}$ (Fig. 4b), where I is the identity matrix. We visualize the operations on the Bloch sphere in Fig. 4c, d. The operation NOT · H corresponds to a π -rotation about the H axis (the line bisecting the z and $-x$ axes), followed by a π -rotation about the x axis. This is equivalent to a rotation about the y axis by $\pi/2$. The operation H · NOT corresponds to a rotation in the opposite direction.

In general, by concatenation of two geometric operations, rotations about arbitrary axes corresponding to a representation of the complete SU(2) group can be realized¹⁴. Applying the scheme presented here—or, alternatively, cavity-induced geometric phase shifts²⁸ to two coupled three-level systems—will complete the universal set of geometric quantum gates and allow for the execution of all-geometric quantum algorithms, which are potentially resilient to noise when short pulses are used²⁹. Moreover, holonomic gates demonstrated for superconducting quantum devices could also be applied to other three-level systems with similar energy level structure.

METHODS SUMMARY

To characterize the gates, we perform full process tomography on the three-level system and reconstruct the process matrix, χ_{exp} , using a maximum-likelihood procedure³⁰. In the presence of dissipative processes, the final density matrix, ρ_f ,

is given by the quantum dynamical map $\rho_f = \sum_{m,n} \chi_{mn} E_m \rho_i E_n^\dagger$ acting on the initial state ρ_i . The full set of nine orthogonal basis operators $\{E_i\} \in \text{SU}(3)$ is chosen as $\{E_i\} = \{I_{01}, \sigma_{01}^x, -i\sigma_{01}^y, \sigma_{01}^z, \sigma_{0e}^x, -i\sigma_{0e}^y, \sigma_{0e}^z, -i\sigma_{1e}^x, I_{1e}\}$, where σ_{ij} are Pauli operators acting on the levels i and j , $I_{01} = |0\rangle\langle 0| + |1\rangle\langle 1|$ and $I_e = |e\rangle\langle e|$. The process is fully determined by its action on a complete set of nine input states, $\{|0\rangle, |e\rangle, |1\rangle, (|0\rangle + |e\rangle)/\sqrt{2}, (|0\rangle + i|e\rangle)/\sqrt{2}, (|0\rangle + |1\rangle)/\sqrt{2}, (|0\rangle + i|1\rangle)/\sqrt{2}, (|e\rangle + |1\rangle)/\sqrt{2}, (|e\rangle + i|1\rangle)/\sqrt{2}\}$. These states are prepared by sequentially applying the identity I , the π -rotation $R_x(\pi)$ and the $\pi/2$ -rotations $R_x(\pi/2)$ and $R_y(\pi/2)$ to the $|0\rangle \leftrightarrow |e\rangle$ and $|e\rangle \leftrightarrow |1\rangle$ transitions. After applying the geometric operation to the input states, we perform full state tomography on the respective output states³¹. The length of a typical sequence is 208 ns (five 40-ns pulses with 2-ns separation). We calibrate the π -pulses on the $|0\rangle \leftrightarrow |e\rangle$ and $|e\rangle \leftrightarrow |1\rangle$ transitions by measuring Rabi oscillations between the corresponding states. From the recorded $9^2 = 81$ measurements, we reconstruct the process matrix, χ_{exp} . The process is compared to the ideal one in equation (1) by calculating its fidelity as $F = \text{Tr}(\chi_{\text{exp}}\chi_{\text{th}})$, where χ_{th} is the process matrix of the ideal process. Because the state $|e\rangle$ serves only as an auxiliary state, we present only the reduced process matrix, $\tilde{\chi}$, which describes the processes involving the states $|0\rangle$ and $|1\rangle$ and omits any operators acting on the state $|e\rangle$. To allow comparison with processes acting on a two-level system, we renormalize $\tilde{\chi}$ by a factor of 3/2. The set of basis operators is thus given by $\{I, X, Y, Z\} = \{I_{01}, \sigma_{01}^x, -i\sigma_{01}^y, \sigma_{01}^z\}$.

Received 16 December 2012; accepted 13 February 2013.

Published online 17 April 2013.

- Pancharatnam, S. Generalized theory of interference and its applications. *Proc. Indian Acad. Sci.* **A44**, 247–262 (1956).
- Berry, M. V. Quantal phase-factors accompanying adiabatic changes. *Proc. R. Soc. Lond. A* **392**, 45–57 (1984).
- Aharonov, Y. & Anandan, J. Phase-change during a cyclic quantum evolution. *Phys. Rev. Lett.* **58**, 1593–1596 (1987).
- Wilczek, F. & Zee, A. Appearance of gauge structure in simple dynamical systems. *Phys. Rev. Lett.* **52**, 2111–2114 (1984).
- Dalibard, J., Gerbier, F., Juzeliūnas, G. & Öhberg, P. Artificial gauge potentials for neutral atoms. *Rev. Mod. Phys.* **83**, 1523–1543 (2011).
- Nayak, C., Simon, S. H., Stern, A., Freedman, M. & Das Sarma, S. Non-Abelian anyons and topological quantum computation. *Rev. Mod. Phys.* **80**, 1083–1159 (2008).
- Pachos, J. K. *Introduction to Topological Quantum Computation* (Cambridge Univ. Press, 2012).
- Zanardi, P. & Rasetti, M. Holonomic quantum computation. *Phys. Lett. A* **264**, 94–99 (1999).
- Duan, L. M., Cirac, J. I. & Zoller, P. Geometric manipulation of trapped ions for quantum computation. *Science* **292**, 1695–1697 (2001).
- Berry, M. Geometric phase memories. *Nature Phys.* **6**, 148–150 (2010).
- Tycko, R. Adiabatic rotational splittings and Berry's phase in nuclear-quadrupole resonance. *Phys. Rev. Lett.* **58**, 2281–2284 (1987).
- Zwanziger, J. W., Koenig, M. & Pines, A. Non-abelian effects in a quadrupole system rotating around two axes. *Phys. Rev. A* **42**, 3107–3110 (1990).
- Anandan, J. Non-adiabatic non-abelian geometric phase. *Phys. Lett. A* **133**, 171–175 (1988).
- Sjöqvist, E. et al. Non-adiabatic holonomic quantum computation. *N. J. Phys.* **14**, 103035 (2012).
- Koch, J. et al. Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A* **76**, 042319 (2007).
- Sjöqvist, E. A new phase in quantum computation. *Physics* **1**, 35 (2008).
- Leek, P. J. et al. Observation of Berry's phase in a solid-state qubit. *Science* **318**, 1889–1892 (2007).
- Möttönen, M., Vartiainen, J. J. & Pekola, J. P. Experimental determination of the Berry phase in a superconducting charge pump. *Phys. Rev. Lett.* **100**, 177201 (2008).
- Filipp, S. et al. Experimental demonstration of the stability of Berry's phase for a spin-1/2 particle. *Phys. Rev. Lett.* **102**, 030404 (2009).
- Wu, H. et al. Geometric phase gates with adiabatic control in electron spin resonance. *Phys. Rev. A* **87**, 032326 (2013).
- Faoro, L., Siewert, J. & Fazio, R. Non-abelian holonomies, charge pumping, and quantum computation with Josephson junctions. *Phys. Rev. Lett.* **90**, 028301 (2003).
- Kamleitner, I., Solinas, P., Müller, C., Shnirman, A. & Möttönen, M. Geometric quantum gates with superconducting qubits. *Phys. Rev. B* **83**, 214518 (2011).
- Solinas, P., Zanardi, P., Zanghi, N. & Rossi, F. Holonomic quantum gates: a semiconductor-based implementation. *Phys. Rev. A* **67**, 062315 (2003).
- Zhu, S. L. & Wang, Z. D. Implementation of universal quantum gates based on nonadiabatic geometric phases. *Phys. Rev. Lett.* **89**, 097902 (2002).
- Paik, H. et al. Observation of high coherence in Josephson junction qubits measured in a three-dimensional circuit QED architecture. *Phys. Rev. Lett.* **107**, 240501 (2011).
- Bianchetti, R. et al. Dynamics of dispersive single-qubit readout in circuit quantum electrodynamics. *Phys. Rev. A* **80**, 043840 (2009).
- Motzoi, F., Gambetta, J. M., Rebentrost, P. & Wilhelm, F. K. Simple pulses for elimination of leakage in weakly nonlinear qubits. *Phys. Rev. Lett.* **103**, 110501 (2009).

28. Pechal, M. *et al.* Geometric phase and nonadiabatic effects in an electronic harmonic oscillator. *Phys. Rev. Lett.* **108**, 170401 (2012).
29. Johansson, M. *et al.* Robustness of non-adiabatic holonomic gates. *Phys. Rev. A* **86**, 062322 (2012).
30. Ježek, M., Fiurášek, J. & Hradil, Z. Quantum inference of states and processes. *Phys. Rev. A* **68**, 012305 (2003).
31. Bianchetti, R. *et al.* Control and tomography of a three level superconducting artificial atom. *Phys. Rev. Lett.* **105**, 223601 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to E. Sjöqvist and members of the EU-funded project GEOMDISS for discussions. We acknowledge feedback on the manuscript from

P. Zanardi and I. Cirac. This work is supported financially by GEOMDISS, the Swiss National Science Foundation and ETH Zurich.

Author Contributions A.A.A. and S.F. developed the idea for the experiment; A.A.A. performed the measurements and analysed the data; J.M.F. designed and fabricated the sample, J.M.F. and K.J. set up the experimental hardware and characterized the sample, M.P. and S.B. contributed to the experiment; A.A.A. and S.F. wrote the manuscript; A.W. and S.F. planned the project; all authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.A.A. (abdumalikov@phys.ethz.ch).

Supplementary Methods); that are included as regulators in the network and have a differential expression score >1 ; or that are strongly differentially expressed (differential expression score = 4); (2) it must include at least 10 representatives from each cluster of genes that have similar expression profiles (using clustering method (4) above); (3) it must contain at least 5 representatives from the predicted targets of each transcription factor in the different networks; (4) it must include a minimal number of representatives from each enriched Gene Ontology (GO) category (computed across all differentially expressed genes); and (5) it must include a manually assembled list of ~ 100 genes that are related to the differentiation process, including the differentially expressed cytokines, cytokine receptors and other cell surface molecules. Because these different criteria might generate substantial overlaps, we used a set-cover algorithm to find the smallest subset of genes that satisfies all of five conditions. We added to this list 18 genes whose expression showed no change (in time or between treatments) in the microarray data.

The 86-gene signature (used for the Fluidigm BioMark qPCR assay) is a subset of the 275-gene signature, selected to include all the key regulators and cytokines discussed. We added to this list 10 control genes (Supplementary Table 5).

Selection of perturbation targets. We used an unbiased approach to rank candidate regulators—transcription factor or chromatin modifier genes—of T_H17 differentiation. Our ranking was based on the following features: (a) whether the gene encoding the regulator belonged to the T_H17 microarray signature (comparing to other $CD4^+$ T cells⁵¹, see Supplementary Methods); (b) whether the regulator was predicted to target key T_H17 molecules (IL-17, IL-21, IL-23r and ROR- γ t); (c) whether the regulator was detected based on both perturbation and physical binding data from the IPA software (<http://www.ingenuity.com/>); (d) whether the regulator was included in the network using a cutoff of at least 10 target genes; (e) whether the gene coding for the regulator was significantly induced in the T_H17 time course—we only consider cases where the induction happened after 4 h to exclude nonspecific hits; (f) whether the gene encoding the regulator was differentially expressed in response to T_H17 -related perturbations in previous studies. For this criterion, we assembled a database of transcriptional effects in perturbed T_H17 cells, including: knockouts of *Batf* (ref. 56), *Rorc* (S. Xiao *et al.*, unpublished), *Hif1a* (ref. 57), *Stat3* and *Stat5* (refs 43, 62), *Tbx21* (A. Awasthi *et al.*, unpublished), *Il23r* (this study), and *Ahr* (ref. 59). We also included data from the T_H17 response to digoxin⁶³ and halofuginone⁶⁴, as well as information on direct binding by ROR- γ t as inferred from ChIP-seq data (S. Xiao *et al.*, unpublished). For each regulator, we counted the number of conditions in which it came up as a significant hit (up/downregulated or bound); for regulators with 2 to 3 hits (quantiles 3 to 7 out of 10 bins), we then assigned a score of 1; for regulators with more than 3 hits (quantiles 8–10), we assigned a score of 2 (a score of 0 is assigned otherwise); and, (g) the differential expression score of the gene in the T_H17 time course.

We ordered the regulators lexicographically by the above features according to the order: (a), (b), (c), (d), (sum of (e) and (f)), (g); that is, first sort according to (a) then break ties according to (b), and so on. We exclude genes that are not over-expressed during at least one time point. As an exception, we retained predicted regulators (features (c) and (d)) that had additional external validation (feature (f)). To validate this ranking, we used a supervised test: we manually annotated 72 regulators that were previously associated with T_H17 differentiation. All of the features are highly specific for these regulators ($P < 10^{-3}$). Moreover, using a supervised learning method (Naive Bayes), the features provided good predictive ability for the annotated regulators (accuracy of 71%, using fivefold cross validation), and the resulting ranking was highly correlated with our unsupervised lexicographic ordering (Spearman correlation >0.86).

We adapted this strategy for ranking protein receptors. To this end, we excluded feature (c) and replaced the remaining 'protein-level' features ((b) and (d)) with the following definitions: (b) whether the respective ligand is induced during the T_H17 time course; and, (d) whether the receptor was included as a target in the network using a cutoff of at least 5 targeting transcriptional regulators.

Gene knockdown using silicon nanowires. 4×4 mm silicon nanowire substrates were prepared and coated with 3 μ l of a 50 μ M pool of four siGENOME siRNAs (Dharmacon) in 96-well tissue culture plates, as previously described¹⁰. Briefly, 150,000 naive T cells were seeded on siRNA-laced nanowires in 10 μ l of complete media and placed in a cell culture incubator (37 °C, 5% CO₂) to settle for 45 min before full media addition. These samples were left undisturbed for 24 h to allow target transcript knockdown. Afterward, siRNA-transfected T cells were activated with anti-CD3/CD28 dynabeads (Invitrogen), according to the manufacturer's recommendations, under T_H17 polarization conditions (TGF- β 1 and IL-6, as above). 10 or 48 h post-activation, culture media was removed from each well and samples were gently washed with 100 μ l of PBS before being lysed in 20 μ l of buffer TCL (Qiagen) supplemented with 2-mercaptoethanol (1:100 by volume).

After mRNA was collected in Turbocapture plates (Qiagen) and converted to cDNA using Sensiscript RT enzyme (Qiagen), qRT-PCR was used to validate both knockdown levels and phenotypic changes relative to 8–12 non-targeting siRNA control samples, as previously described⁶⁵. A 60% reduction in target mRNA was used as the knockdown threshold. In each knockdown experiment, each individual siRNA pool was run in quadruplicate; each siRNA was tested in at least three separate experiments (Supplementary Fig. 9).

mRNA measurements in perturbation assays. We used the nCounter system, presented in full in ref. 66, to measure a custom CodeSet constructed to detect a total of 293 genes, selected as described above. We also used the Fluidigm BioMark HD system to measure a smaller set of 96 genes. Finally, we used RNA-seq to follow up and validate 12 of the perturbations. Details of the experimental and analytical procedures of these analyses are provided in the Supplementary Methods.

Profiling Tsc22d3 DNA binding using ChIP-seq. ChIP-seq for Tsc22d3 was performed as previously described⁶⁷ using an antibody from Abcam. The analysis of this data was performed as previously described⁷ and is detailed in the Supplementary Methods.

Estimating statistical significance of monochromatic interactions between modules. The functional network in Fig. 4b consists of two modules: positive and negative. We compute two indices: (1) within-module index: the percentage of positive edges between members of the same module (that is, downregulation in knockdown/knockout); and (2) between-module index: the percentage of negative edges between members of different modules. We shuffled the network 1,000 times, while maintaining the nodes' out degrees (that is, number of outgoing edges) and edges' signs (positive/negative), and re-computed the two indices. The reported *P* values were computed using a *t*-test.

46. Kim, U. *et al.* The B-cell-specific transcription coactivator OCA-B/OBF-1/Bob-1 is essential for normal production of immunoglobulin isotypes. *Nature* **383**, 542–547 (1996).
47. Wang, V. E., Tantin, D., Chen, J. & Sharp, P. A. B cell development and immunoglobulin transcription in Oct-1-deficient mice. *Proc. Natl Acad. Sci. USA* **101**, 2005–2010 (2004).
48. Bettelli, E. *et al.* Reciprocal developmental pathways for the generation of pathogenic effector T_H17 and regulatory T cells. *Nature* **441**, 235–238 (2006).
49. Awasthi, A. *et al.* A dominant function for interleukin 27 in generating interleukin 10-producing anti-inflammatory T cells. *Nature Immunol.* **8**, 1380–1389 (2007).
50. Awasthi, A. *et al.* Cutting edge: IL-23 receptor gfp reporter mice reveal distinct populations of IL-17-producing cells. *J. Immunol.* **182**, 5904–5908 (2009).
51. Wei, G. *et al.* Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating $CD4^+$ T cells. *Immunity* **30**, 155–167 (2009).
52. Reich, M. *et al.* GenePattern 2.0. *Nature Genet.* **38**, 500–501 (2006).
53. Storey, J., Xiao, W., Leek, J., Tompkins, R. & Davis, R. Significance analysis of time course microarray experiments. *Proc. Natl Acad. Sci. USA* **102**, 12837–12842 (2005).
54. Leek, J. T., Monsen, E., Dabney, A. R. & Storey, J. D. EDGE: extraction and analysis of differential gene expression. *Bioinformatics* **22**, 507–508 (2006).
55. Chechik, G. & Koller, D. Timing of gene expression responses to environmental changes. *J. Comput. Biol.* **16**, 279–290 (2009).
56. Schraml, B. U. *et al.* The AP-1 transcription factor Batf controls T_H17 differentiation. *Nature* **460**, 405–409 (2009).
57. Shi, L. Z. *et al.* HIF1 α -dependent glycolytic pathway orchestrates a metabolic checkpoint for the differentiation of T_H17 and T_{reg} cells. *J. Exp. Med.* **208**, 1367–1376 (2011).
58. Durant, L. *et al.* Diverse targets of the transcription factor STAT3 contribute to T cell pathogenicity and homeostasis. *Immunity* **32**, 605–615 (2010).
59. Jux, B., Kadow, S. & Esser, C. Langerhans cell maturation and contact hypersensitivity are impaired in aryl hydrocarbon receptor-null mice. *J. Immunol.* **182**, 6709–6717 (2009).
60. Xiao, S. *et al.* Retinoic acid increases Foxp3⁺ regulatory T cells and inhibits development of Th17 cells by enhancing TGF- β -driven Smad3 signaling and inhibiting IL-6 and IL-23 receptor expression. *J. Immunol.* **181**, 2277–2284 (2008).
61. Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* **112**, 531–552 (1999).
62. Durant, L. *et al.* Diverse targets of the transcription factor STAT3 contribute to T cell pathogenicity and homeostasis. *Immunity* **32**, 605–615 (2010).
63. Huh, J. R. *et al.* Digoxin and its derivatives suppress T_H17 cell differentiation by antagonizing ROR γ t activity. *Nature* **472**, 486–490 (2011).
64. Sundrud, M. S. *et al.* Halofuginone inhibits T_H17 cell differentiation by activating the amino acid starvation response. *Science* **324**, 1334–1338 (2009).
65. Chevrier, N. *et al.* Systematic discovery of TLR signaling components delineates viral-sensing circuits. *Cell* **147**, 853–867 (2011).
66. Geiss, G. K. *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnol.* **26**, 317–325 (2008).
67. Ram, O. *et al.* Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell* **147**, 1628–1639 (2011).

Optical magnetic imaging of living cells

D. Le Sage^{1,2*}, K. Arai^{3*}, D. R. Glenn^{1,2,4*}, S. J. DeVience⁵, L. M. Pham⁶, L. Rahn-Lee⁷, M. D. Lukin², A. Yacoby², A. Komeili⁷ & R. L. Walsworth^{1,2,4}

Magnetic imaging is a powerful tool for probing biological and physical systems. However, existing techniques either have poor spatial resolution compared to optical microscopy and are hence not generally applicable to imaging of sub-cellular structure (for example, magnetic resonance imaging¹), or entail operating conditions that preclude application to living biological samples while providing submicrometre resolution (for example, scanning superconducting quantum interference device microscopy², electron holography³ and magnetic resonance force microscopy⁴). Here we demonstrate magnetic imaging of living cells (magnetotactic bacteria) under ambient laboratory conditions and with sub-cellular spatial resolution (400 nanometres), using an optically detected magnetic field imaging array consisting of a nanometre-scale layer of nitrogen-vacancy colour centres implanted at the surface of a diamond chip. With the bacteria placed on the diamond surface, we optically probe the nitrogen-vacancy quantum spin states and rapidly reconstruct images of the vector components of the magnetic field created by chains of magnetic nanoparticles (magnetosomes) produced in the bacteria. We also spatially correlate these magnetic field maps with optical images acquired in the same apparatus. Wide-field microscopy allows parallel optical and magnetic imaging of multiple cells in a population with submicrometre resolution and a field of view in excess of 100 micrometres. Scanning electron microscope images of the bacteria confirm that the correlated optical and magnetic images can be used to locate and characterize the magnetosomes in each bacterium. Our results provide a new capability for imaging bio-magnetic structures in living cells under ambient conditions with high spatial resolution, and will enable the mapping of a wide range of magnetic signals within cells and cellular networks^{5,6}.

Nitrogen-vacancy (NV) colour centres in diamond (see Methods for details) enable nanometre-scale magnetic sensing and imaging under ambient conditions^{7,8}. As recently shown using a variety of methods^{9,10}, NV centres within room-temperature diamond can be brought into close proximity (a few nanometres) of magnetic field sources of interest while maintaining long NV electronic spin coherence times (of the order of milliseconds), a large (about one Bohr magneton) Zeeman shift of the NV spin states, and optical preparation and readout of the NV spin. Recent demonstrations of NV-diamond magnetometry include high-precision sensing and submicrometre imaging of externally applied and controlled magnetic fields^{6,9–11}; detection of electron¹² and nuclear^{13–15} spins; and imaging of a single electron spin within a neighbouring diamond crystal with ~ 10 nm resolution¹⁶. However, a key challenge for NV-diamond magnetometry is submicrometre imaging of spins and magnetic nanoparticles located outside the diamond crystal and within a target of interest. Here we present the first such demonstration of NV-diamond imaging of the magnetic field distribution produced by a living biological specimen.

Magnetotactic bacteria (MTB) are of considerable interest as a model system for the study of molecular mechanisms of biomineralization^{17,18} and have often been used for testing novel biomagnetic imaging

modalities^{3,19–21}. MTB form magnetosomes, membrane-bound organelles containing nanoparticles of magnetite (Fe_3O_4) or greigite (Fe_3S_4), that are arranged in chains with a net dipole moment, allowing the bacteria to orient and travel along geomagnetic field lines (magnetotaxis)^{17,18}. Magnetic nanoparticles produced in the magnetosomes are chemically pure, single-domain monocrystalline ferrimagnets, with species-specific morphologies and strikingly uniform size distributions^{17,18}. These features, combined with the ease of biofunctionalization and aqueous dispersion afforded by the magnetosome membrane²², make synthesis of magnetic nanoparticles by MTB an attractive research area for various biomedical applications^{18,22}, including magnetic labelling, separation and drug delivery, as well as local hyperthermic cancer treatment and contrast enhancement in magnetic resonance imaging. For the NV-diamond bio-magnetic imaging demonstrations presented here (see Fig. 1), we used *Magnetospirillum magneticum* AMB-1, an MTB strain that forms magnetic nanoparticles with cubo-octahedral morphology and an average diameter of ~ 50 nm. (Figure 1c shows a

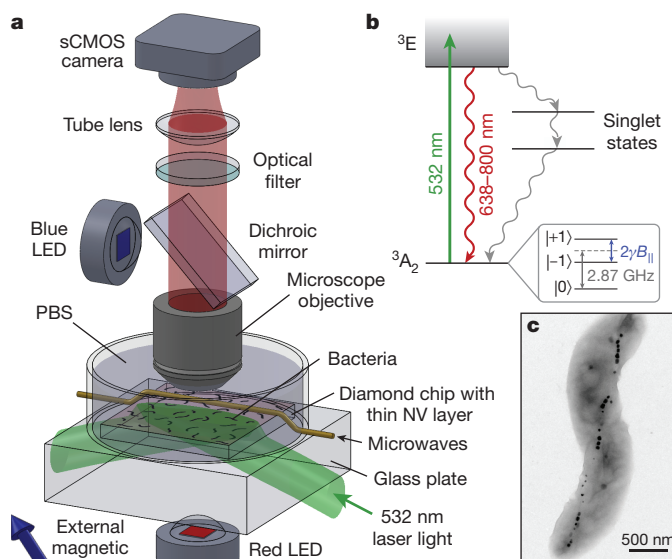


Figure 1 | Wide-field magnetic imaging microscope. **a**, Custom-built wide-field fluorescence microscope used for combined optical and magnetic imaging. Live magnetotactic bacteria (MTB) are placed in phosphate-buffered saline (PBS) on the surface of a diamond chip implanted with nitrogen-vacancy (NV) centres. Vector magnetic field images are derived from optically detected magnetic resonance (ODMR)^{9–10} interrogation of NV centres excited by a totally-internally-reflected 532 nm laser beam, and spatially correlated with bright field optical images. See text for details. LED, light-emitting diode. **b**, Energy-level diagram of the NV centre; see Methods for details. **c**, Typical transmission electron microscope (TEM) image of an *M. magnetotactic* AMB-1 bacterium. Magnetite nanoparticles appear as spots of high electron density.

¹Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA. ²Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA. ³Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁴Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA. ⁵Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ⁶School of Engineering and Applied Science, Harvard University, Cambridge, Massachusetts 02138, USA. ⁷Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, California 94720, USA.

*These authors contributed equally to this work.

transmission electron microscopy image exhibiting the characteristic morphology of *M. magneticum* AMB-1, including a chain of magnetic nanoparticles distributed over the length of the cell. Gaps between nanoparticles are common in AMB-1 (ref. 23).)

We acquired correlated magnetic field and optical images of populations of MTB using the NV-diamond wide-field imager depicted schematically in Fig. 1a (ref. 6). The system was operated in two distinct configurations, one optimized for rapid magnetic imaging of living cells in a liquid medium, and the other for high-precision measurements of stable magnetic field patterns produced by dry bacteria on the diamond surface. In both cases, magnetic imaging was carried out using a pure diamond chip doped with a 10-nm-deep surface layer of NV centres. NV electronic spin states were optically polarized and interrogated with green illumination (wavelength $\lambda = 532$ nm), coherently manipulated using resonant microwave fields, and detected via spin-state-dependent fluorescence in the red (Fig. 1b). NV electronic spin resonance frequencies are Zeeman-shifted in the presence of a local external magnetic field (such as from magnetic nanoparticles in an MTB), allowing NV-fluorescence-based magnetometry by optically detected magnetic resonance (ODMR)^{8–10}. Four independent ODMR measurements enabled determination of all vector components of the magnetic field within each imaging pixel (see Methods). For imaging of live samples, the green excitation beam was directed into the diamond chip at an angle greater than the critical angle for the diamond–water interface, resulting in total internal reflection of high-intensity green light within the diamond, while low-intensity red NV fluorescence passed freely to the objective and was imaged onto the sCMOS (scientific complementary metal-oxide semiconductor) camera (Fig. 1a). Cells at the diamond surface were thereby decoupled from high optical intensity, allowing NV magnetic imaging times up to several minutes while maintaining cellular viability. For magnetic imaging of dry bacteria, the green excitation beam could be configured in the same manner as for live/wet samples, or be allowed to pass directly through the sample, normal to the diamond surface, with comparable optical and magnetic imaging results.

We obtained optical images of the magnetic field distributions produced by multiple cells on the diamond surface across a wide field of view ($100\ \mu\text{m} \times 30\ \mu\text{m}$) and with high spatial resolution (~ 400 nm) using a sCMOS camera (Fig. 2). We concurrently acquired bright-field optical images using red ($\lambda = 660$ nm) LED illumination to enable correlation of cell positions and morphology with the observed magnetic field patterns. Immediately following magnetic imaging, the MTB were stained and imaged in fluorescence under blue ($\lambda = 470$ nm) LED excitation to perform a bacterial viability assay (see Methods), using a conservative viability threshold that excluded non-viable bacteria with 99% certainty (see Supplementary Methods). Under appropriate imaging conditions, the magnetic field patterns produced by the MTB could be measured within 4 min with minimal cellular radiation exposure, such that a significant fraction of the MTB remained alive after magnetic and bright-field imaging. For example, $\sim 44\%$ of the MTB in the field of view shown in Fig. 2a, b were found to be viable after magnetic and bright-field imaging, compared to 54% viability for cells directly from culture. Many of these living MTB produced magnetic field signals with large signal-to-noise ratios (~ 10). For high-precision characterization of the bacterial magnetic fields and comparison to electron microscope images, we also carried out a series of measurements using dried MTB samples on the diamond surface, imaged using a high-numerical-aperture (high-NA) air objective (Fig. 2c, d). Relaxing the requirement of maintaining cellular viability allowed for longer magnetic image averaging times, with concomitant reduction in photon shot-noise. Also, elimination of both the poly-L-lysine adhesion layer (see Methods) and residual cellular Brownian motion in liquid brought the cells closer to the diamond substrate and improved their spatial stability, resulting in higher time-averaged magnetic fields at the layer of NV centres near the diamond surface. We thus expect that the dried cell technique may be the preferred approach for biological

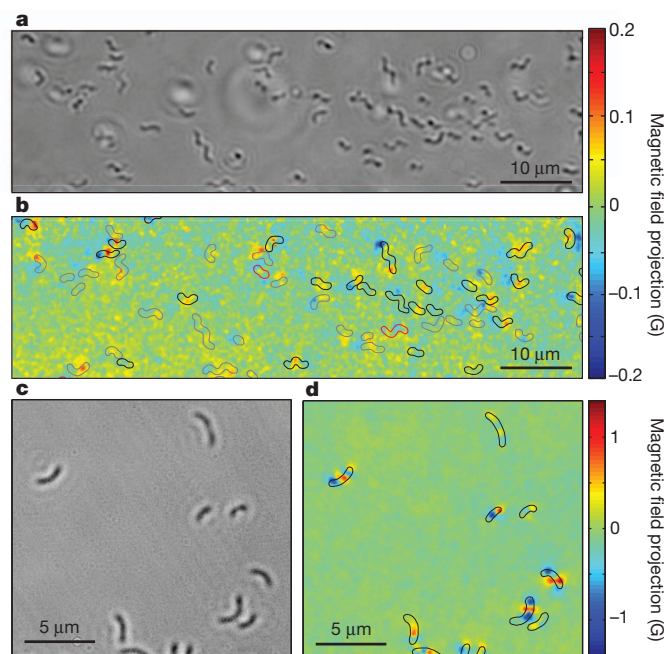


Figure 2 | Wide-field optical and magnetic images of magnetotactic bacteria. **a**, Bright-field optical image of MTB adhered to the diamond surface while immersed in PBS. **b**, Image of magnetic field projection along the [111] crystallographic axis in the diamond for the same region as **a**, determined from NV ODMR. Superimposed outlines indicate MTB locations determined from **a**. Outline colours indicate results of the live-dead assay performed after measuring the magnetic field (black for living, red for dead, and grey for indeterminate). **c**, Bright-field image of dried MTB on the diamond chip. **d**, Image of magnetic field projection along [111] for the same region, with outlines indicating MTB locations determined from **c**.

applications that do not require sustained imaging of magnetic fields produced by developing cells.

As shown in Figs 2–4, the NV-diamond wide-field imager enables rapid, simultaneous measurement of biomagnetic particle distributions in many MTB, with magnetic field sensitivity and spatial resolution sufficient both to localize magnetic nanoparticles within individual MTB and to quantify the MTB magnetic moment from the magnetic field images. To verify these capabilities, we recorded scanning electron microscope (SEM) images of dried MTB in place on the surface of the diamond chip after the magnetic and bright-field imaging had been completed. Positions and relative sizes of the magnetic nanoparticles within each MTB were determined from the backscattered electron SEM images, and used to calculate the expected vector magnetic field pattern from the MTB (up to a normalization constant equivalent to the total magnetic moment of the particles—see Methods). The magnetic field patterns that we calculated (from SEM data) and measured (with the NV-diamond imager) were in excellent agreement (Fig. 3a–h), across a wide variety of magnetic nanoparticle distributions within the MTB (Fig. 4). We also determined the total magnetic moment of each MTB (for example, $(1.2 \pm 0.1) \times 10^{-16}$ A m² for the MTB in Fig. 3a–h) by numerically fitting the modelled field distribution to the measured distribution, leaving the standoff distance and magnetic moment as free parameters. From such optical magnetic field measurements, we determined the distribution of magnetic moments from 36 randomly-sampled MTB on the diamond surface (Fig. 3i), with a mean value (0.5×10^{-16} A m²) that was consistent with previous estimates of the average moment per MTB for *M. magneticum* AMB-1 (ref. 24), although our measurements showed that most AMB-1 cells had smaller moments. Note that most previously applied magnetic measurement techniques determine the average properties of large MTB populations^{24,25} but are insensitive to variations among individuals within the population. In contrast, the ability of the NV-diamond

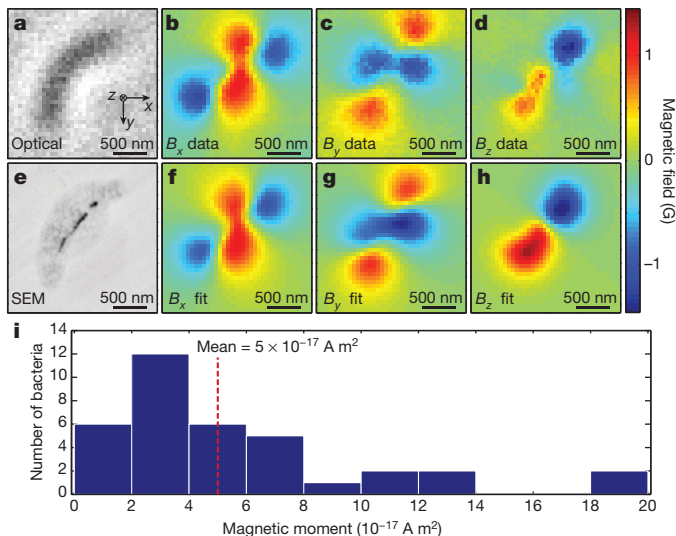


Figure 3 | Determining magnetic moments of individual bacteria from measured magnetic field distributions. **a**, Bright-field image of an MTB. **b–d**, Measured magnetic field projections along the *x* axis (B_x ; **b**), *y* axis (B_y ; **c**) and *z* axis (B_z ; **d**) within the same field-of-view. **e**, Scanning electron microscope (SEM) image of the same bacterium. **f–h**, Simulated magnetic field projections along the *x* axis (**f**), *y* axis (**g**) and *z* axis (**h**), assuming that magnetic nanoparticle locations match those extracted from **e**. The total magnetic moment was determined from the best fit of the calculated field distribution to the measurement (see Methods for details). **i**, Magnetic moments of 36 randomly-sampled MTB, as determined from optical magnetic field images and modelled field distributions.

wide-field magnetic imager to measure rapidly the magnetic properties of many individuals in an MTB population provides a robust tool to investigate the defects of various biomineralization mutants, making it possible to distinguish between defects that equally affect all cells in a population versus those that disproportionately disrupt magnetosome formation in a subset of cells. The *M. magneticum* AMB-1 bacteria studied here provided high signal-to-noise ratio magnetic imaging

data, even though the typical magnetic moments of these bacteria are an order of magnitude smaller than many commonly studied MTB strains^{3,21}. This suggests that NV magnetic imaging will be applicable to a broad variety of MTB.

Furthermore, we were able to determine the positions of magnetic nanoparticle chains in individual MTB from the magnetic field distributions measured with the NV-diamond imager, even without the use of correlated SEM data, by noting that the magnetic nanoparticle chain endpoints occurred at locations of maximum field divergence (yellow bars in Fig. 4). Distinct groups of magnetic nanoparticles could be resolved if their separation was more than the 400 nm diffraction-limited resolution of our optical magnetometry measurements (for example, Fig. 4d), and endpoints of single, well-isolated magnetic nanoparticle chains could be localized to within <100 nm (for example, Fig. 4b). Using the chain positions and a simplified model for the magnetic nanoparticle field-source distribution, we estimated the total magnetic moments of individual MTB from the magnetic field data alone (without correlated SEM measurements). The magnetic moments determined using this analysis procedure (for example, $0.9 \times 10^{-16} \text{ A m}^2$ for the MTB in Figs 3a–h and 4a, using the estimated chain position in Fig. 4a) agreed well with the values derived using the more detailed SEM-based models when the magnetic nanoparticles were arranged in long chains.

The NV-diamond wide-field imager provides powerful new capabilities that could shed light on unanswered questions regarding the development of MTB magnetic properties^{17,18}. Some existing methods can probe the internal magnetic structure of a single MTB^{3,19}, or measure the magnetic field²⁰ or field gradient²¹ near a single MTB, but only NV magnetic imaging provides direct magnetic field measurements with sub-cellular resolution under ambient environmental conditions—opening the way to real-time imaging of magnetic nanoparticle formation and chain dynamics in single living MTB. Real-time magnetic measurements will enable observation of the transition of magnetic nanoparticles from superparamagnetic to permanent, single-magnetic-domain states as the nanoparticles grow¹⁸. The ability to locate chains of nanoparticles from the magnetic images will make it possible to measure the movement of magnetosome chains across the cell-division cycle of individual MTB.

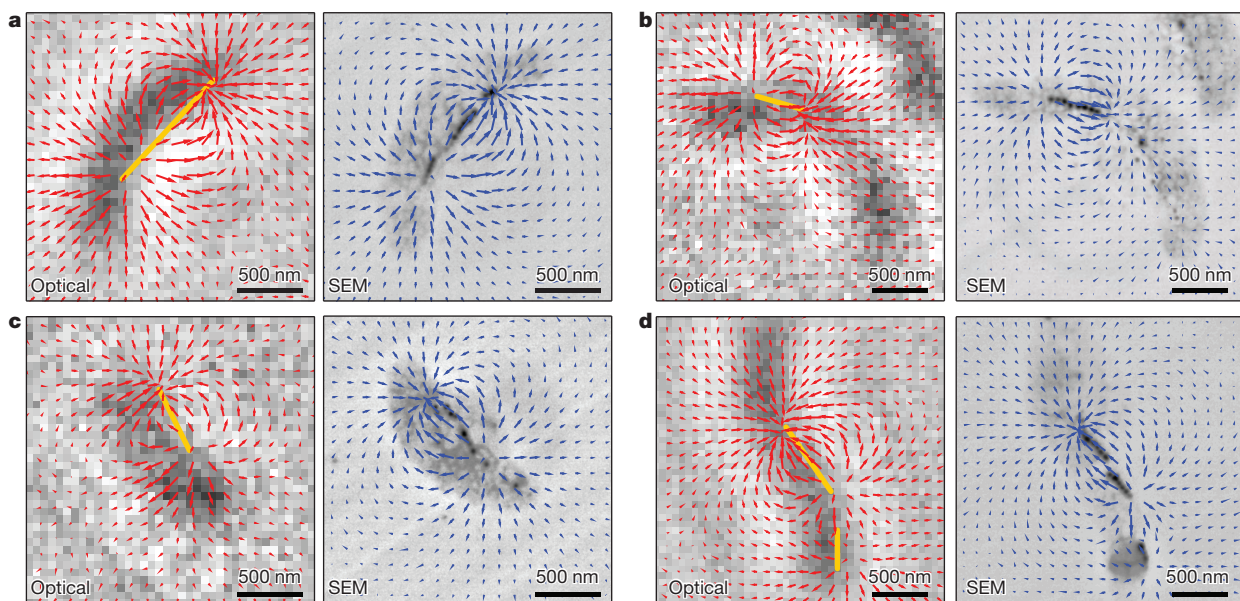


Figure 4 | Localization of magnetic nanoparticle chains using magnetic field measurements. **a**, Vector plots of the measured (red arrows, left panel) and simulated (blue arrows, right panel) magnetic field projections in the *x–y* plane, for the same MTB as in Fig. 3a–h, superimposed on the optical and backscattered electron images, respectively. The estimated location of the

magnetic nanoparticle chain inside the MTB (yellow bar, left panel), as determined from the divergence of the measured magnetic field, coincides well with the magnetic nanoparticle positions found by SEM. **b–d**, The same information as presented in **a**, but for three different MTB. In **d**, two distinct magnetic nanoparticle chains are identified (yellow bars, left panel).

The measurements presented here are also directly applicable to studying the formation of magnetic nanoparticles in other organisms²⁶. Such formation is of interest for MRI contrast enhancement²⁷, and has been linked with neurodegenerative disorders²⁸; it has also been proposed as a mechanism for magnetic navigation in higher organisms^{26,29,30}. In particular, there is great current interest in identifying potential vertebrate magnetoreceptor cells³⁰, which are believed to have a magnetic moment that is comparable to or larger than found in MTB, suggesting that high-throughput NV-diamond magnetic imaging could be a valuable tool for localizing magnetic cells in a broad range of tissue samples. More generally, with further improvements in detector sensitivity and the use of spin-echo techniques for the detection of time-dependent fields^{6,7,11}, NV-diamond magnetic imaging could be applied to a variety of biologically interesting systems, including firing patterns in neuronal cultures^{5,6}, detection of free radicals generated by signalling or immune responses, and the localization of molecules tagged with specific spin labels.

METHODS SUMMARY

Wide-field ODMR measurements were acquired of *M. magneticum* AMB-1 bacteria adhered to a diamond chip with a 10-nm-thick layer of NV centres 10 nm from the surface. For wet samples, a thin layer of poly-L-lysine was deposited on the diamond to improve cellular adhesion. A uniform 37-G external magnetic field was applied to separate the $|\pm 1\rangle$ spin states and to select the NV axis of interest. The magnetic-field shifts along the NV axis were extracted by fitting Lorentzian lines to the ODMR signals from each pixel of the image. For wet samples, a fluorescence-based bacterial viability assay (Molecular Probes BacLight kit) was carried out to determine which cells remained alive after imaging. For samples of dried cells, the magnetic imaging was repeated for all four NV axes to create a two-dimensional image of the magnetic field along all three Cartesian directions; the diamond with bacteria was then imaged with a field emission SEM (Zeiss Sigma) using backscatter mode to identify the locations of the magnetosomes within the bacteria. A nonlinear fit was performed on simulated magnetic field images calculated from the positions and sizes of the magnetosomes to find the standoff distance and magnetic moments of the magnetosome chains. Magnetosome chain locations and directions were also estimated using the measured magnetic field divergence, and the magnetic moment of the chain was calculated by modelling each chain as a continuous row of magnetic dipoles.

Full Methods and any associated references are available in the online version of the paper.

Received 4 February; accepted 14 March 2013.

1. Lee, S.-C. *et al.* MR microscopy of micron scale structures. *Magn. Reson. Imaging* **27**, 828–833 (2009).
2. Finkler, A. *et al.* Self-aligned nanoscale SQUID on a tip. *Nano Lett.* **10**, 1046–1049 (2010).
3. Dunin-Borkowski, R. E. *et al.* Magnetic microstructure of magnetotactic bacteria by electron holography. *Science* **282**, 1868–1870 (1998).
4. Degen, C. L., Poggio, M., Mamin, H. J., Rettner, C. T. & Rugar, D. Nanoscale magnetic resonance imaging. *Proc. Natl Acad. Sci. USA* **106**, 1313–1317 (2009).
5. Hall, L. T. *et al.* Monitoring ion-channel function in real time through quantum decoherence. *Proc. Natl Acad. Sci. USA* **107**, 18777–18782 (2010).
6. Pham, L. M. *et al.* Magnetic field imaging with nitrogen-vacancy ensembles. *N. J. Phys.* **13**, 045021 (2011).
7. Maze, J. R. *et al.* Nanoscale magnetic sensing with an individual electronic spin in diamond. *Nature* **455**, 644–647 (2008).
8. Balasubramanian, G. *et al.* Nanoscale imaging magnetometry with diamond spins under ambient conditions. *Nature* **455**, 648–651 (2008).
9. Maletinsky, P. *et al.* A robust scanning diamond sensor for nanoscale imaging with single nitrogen-vacancy centres. *Nature Nanotechnol.* **7**, 320–324 (2012).

10. Steinert, S. *et al.* High sensitivity magnetic imaging using an array of spins in diamond. *Rev. Sci. Instrum.* **81**, 043705 (2010).
11. Le Sage, D. *et al.* Efficient photon detection from color centers in a diamond optical waveguide. *Phys. Rev. B* **85**, 121202(R) (2012).
12. Hanson, R., Mendoza, F. M., Epstein, R. J. & Awschalom, D. D. Polarization and readout of coupled single spins in diamond. *Phys. Rev. Lett.* **97**, 087601 (2006).
13. Childress, L. *et al.* Coherent dynamics of coupled electron and nuclear spin qubits in diamond. *Science* **314**, 281–285 (2006).
14. Mamin, H. J. *et al.* Nanoscale nuclear magnetic resonance with a nitrogen-vacancy spin sensor. *Science* **339**, 557–560 (2013).
15. Staudacher, T. *et al.* Nuclear magnetic resonance spectroscopy on a (5-nanometer)³ sample volume. *Science* **339**, 561–563 (2013).
16. Grinolds, M. S. *et al.* Nanoscale magnetic imaging of a single electron spin under ambient conditions. *Nature Phys.* **9**, 215–219 (2013).
17. Komeili, A. Molecular mechanisms of compartmentalization and biomineralization in magnetotactic bacteria. *FEMS Microbiol. Rev.* **36**, 232–255 (2012).
18. Faivre, D. & Schüler, D. Magnetotactic bacteria and magnetosomes. *Chem. Rev.* **108**, 4875–4898 (2008).
19. Lam, K. P. *et al.* Characterizing magnetism of individual magnetosomes by X-ray magnetic circular dichroism in a scanning transmission X-ray microscope. *Chem. Geol.* **270**, 110–116 (2010).
20. Qian, L. *et al.* Magnetic characterization of individual magnetotactic bacteria. (APS March Meeting 2011, 2011); <http://meetings.aps.org/link/BAPS.2011.MAR.D16.8> (abstract published online, 2011).
21. Proksch, R. B. *et al.* Magnetic force microscopy of the submicron magnetic assembly in a magnetotactic bacterium. *Appl. Phys. Lett.* **66**, 2582–2584 (1995).
22. Matsunaga, T., Suzuki, T., Tanaka, M. & Arakaki, A. Molecular analysis of magnetotactic bacteria and development of functional bacterial magnetic particles for nano-biotechnology. *Trends Biotechnol.* **25**, 182–188 (2007).
23. Draper, O. *et al.* MamK, a bacterial actin, forms dynamic filaments *in vivo* that are regulated by the acidic proteins MamJ and LimJ. *Mol. Microbiol.* **82**, 342–354 (2011).
24. Krichinsky, A. *et al.* Trapping motile magnetotactic bacteria with a magnetic recording head. *J. Appl. Phys.* **101**, 014701 (2007).
25. Moskowitz, B. M., Frankel, R. B. & Bazylinski, D. A. Rock magnetic criteria for the detection of biogenic magnetite. *Earth Planet. Sci. Lett.* **120**, 283–300 (1993).
26. Pósfai, M. & Dunin-Borkowski, R. E. Magnetic nanocrystals in organisms. *Elements* **5**, 235–240 (2009).
27. Zurkiya, O., Chan, A. W. S. & Hu, X. Mag A is sufficient for producing magnetic nanoparticles in mammalian cells, making it an MRI reporter. *Magn. Reson. Med.* **59**, 1225–1231 (2008).
28. Dobson, J. Magnetic iron compounds in neurological disorders. *Ann. NY Acad. Sci.* **1012**, 183–192 (2004).
29. Mora, C. V., Davison, M., Wild, J. M. & Walker, M. M. Magnetoreception and its trigeminal mediation in the homing pigeon. *Nature* **432**, 508–511 (2004).
30. Eder, S. H. K. *et al.* Magnetic characterization of isolated candidate vertebrate magnetoreceptor cells. *Proc. Natl Acad. Sci. USA* **109**, 12022–12027 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Qian for discussions about related experiments using a scanning SQUID microscope in the laboratory of K. A. Moler, J. W. Lichtman and R. Schalek for access to the SEM, S. G. Turney for access to the water-immersion objective and advice regarding cell immobilization, and P. R. Hemmer and H. Park for technical discussions. A.K. was supported by a David and Lucille Packard Foundation Fellowship in Science and Engineering and by the National Institutes of Health (R01GM084122). This work was supported by the NSF and the DARPA QuASAR programme.

Author Contributions D.L. and R.L.W. conceived the idea of the study. K.A. developed modelling and fitting algorithms to interpret the data. D.L., K.A., D.R.G., S.J.D. and L.M.P. performed magnetic, optical and SEM imaging experiments, and analysed data. L.R.-L. and A.K. provided bacteria cultures and TEM images. M.D.L., R.L.W. and A.Y. conceived the application of the NV-diamond wide-field imager to biomagnetism. All authors discussed the results and participated in writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.L.W. (rwalsworth@cfa.harvard.edu).

METHODS

NV physics. The NV centre consists of a substitutional nitrogen atom adjacent to a vacancy in the diamond lattice (see Supplementary Fig. 1). The NV centre has a spin-triplet ground state with a 2.87 GHz zero-field splitting between the $|0\rangle$ and $|\pm 1\rangle$ spin states (see Fig. 1b). Optical excitation of an NV centre primarily produces a spin-conserving excitation and decay process, resulting in the emission of a photon in the 640–800 nm wavelength band. However, the $|\pm 1\rangle$ excited states also decay non-radiatively about one-third of the time to the $|0\rangle$ ground state via metastable singlet states. This leads to both optical polarization into the $|0\rangle$ ground state and state-dependent fluorescence rates that may be used to optically distinguish the $|0\rangle$ state from the $|\pm 1\rangle$ states.

The magnetic field projection at an NV centre's location can be measured by monitoring the fluorescence rate of the NV centre during continuous optical excitation, while varying the frequency of a continuous microwave drive^{8–10}. When the applied microwave frequency is on resonance with either of the $|0\rangle \leftrightarrow |\pm 1\rangle$ state transitions, some of the NV state population is transferred from the $|0\rangle$ optically-pumped state to a mixed state, and consequently, the fluorescence rate decreases.

The NV centre's zero-field splitting quantizes the spin states along the NV symmetry axis (indicated by a blue rod in Supplementary Fig. 1). Depending upon the relative positions of the nitrogen atom and vacancy, this symmetry axis can lie along one of four possible crystallographic directions within the diamond lattice (other possible crystallographic axes are indicated by yellow rods in Supplementary Fig. 1). In an external magnetic field, the $|0\rangle \leftrightarrow |\pm 1\rangle$ spin-flip transition frequencies shift by $\Delta f = \pm \gamma B_{\parallel}$ (see Fig. 1c), where $\gamma = 2.8 \text{ MHz G}^{-1}$ is the gyro-magnetic ratio of the NV electronic spin, and B_{\parallel} is the magnetic field projection along the NV symmetry axis.

Diamond samples. Magnetic field sensing was carried out using high-purity, single-crystal diamond chips. For imaging wet bacterial samples, we used an electronic-grade diamond (3 mm \times 3 mm \times 0.5 mm) grown using chemical vapour deposition (CVD) by Element Six Ltd. The diamond was implanted with $^{15}\text{N}^+$ ions at 14 keV energy and annealed at 1,200 °C to produce a 10-nm-thick layer of NV centres 20 nm beneath the surface of the diamond (as estimated using Stopping and Range of Ions in Matter (SRIM) software). The estimated NV surface density within the layer was 3×10^{11} NV per cm^2 . For imaging dry bacterial samples, we used a high-purity, single-crystal diamond chip (1.5 mm \times 1.5 mm \times 0.3 mm) manufactured by Sumitomo Electric Industries using the high-pressure, high-temperature (HPHT) method. This diamond was implanted with $^{15}\text{N}^+$ ions with 15 keV energy and then annealed at 800 °C to produce a 10-nm-thick layer of NV centres 10 nm beneath the surface of the diamond (as estimated using SRIM), with an estimated surface density of 1×10^{12} NV per cm^2 .

Wide-field magnetic imaging microscope. NV centres were optically excited with a 532 nm laser (Changchun New Industries) switched on and off by an acousto-optic modulator (Isomet, M1133-aQ80L-1.5). A small fraction of the laser light was split off and directed onto a photodiode (Thorlabs), and the resulting signal was sent to a servo-lock system (New Focus) to amplitude-stabilize the excitation beam using the same acousto-optic modulator. For imaging of bacterial samples in liquid, laser light was coupled into the diamond from below through a polished glass cube (constructed from two right-angle prisms, Thorlabs), to which the diamond was affixed by optical adhesive (Norland). The peak intensity of the totally-internally-reflected laser light at the interior surface of the diamond was measured in this case to be $\sim 1 \text{ kW cm}^{-2}$. We also note that for our angle of incidence at the diamond–water interface, $\theta_{\text{dw}} \approx 39^\circ$, the calculated attenuation length for the evanescent wave intensity is $d_{\text{dw}} = 58 \text{ nm}$. For imaging of dry samples, laser light could be configured in the same manner as for live/wet samples, or directed onto the bacteria from below, normal to the diamond surface. Dry sample data presented here were acquired using the latter method.

A 660-nm-wavelength LED (Thorlabs) was used to back-illuminate the sample for bright-field images. Excitation of fluorescence dyes used in the bacterial viability assays (see below) was carried out with a 470-nm LED (Thorlabs), directed onto the sample through the microscope objective. Optical fluorescence or transmitted red LED light was collected by the objective (Olympus, UIS2 LumFLN 60xW /1.1 NA for wet samples; Olympus, MPlan FLN 100 \times /0.90 NA for dry samples), passed through a dichroic mirror (Thorlabs for wet samples; Semrock for dry samples) and an optical filter (Semrock for NV fluorescence and transmitted red light; emission filters as described below for fluorescence from bacterial viability assay dyes), and imaged onto a digital camera (Andor for wet samples; Starlight Xpress for dry samples). The output of a microwave synthesizer (SRS) was controlled by a switch (Mini-Circuits), then amplified (Mini-Circuits) and applied to the diamond with a wire. A permanent magnet was used to apply a uniform external magnetic field.

ODMR measurements. *M. magneticum* AMB-1 cells were grown statically in 1.5-ml microcentrifuge tubes filled with 1.5 ml of growth medium (described in

ref. 31, but with 0.1 g l^{-1} of sodium thiosulphate). For measurements of wet samples, the diamond surface was prepared by placing a drop ($\sim 5 \mu\text{l}$) of 0.01% poly-L-lysine solution (Sigma molecular mass 70–150 kDa) on its surface, which was then allowed to dry. The bath around the diamond (contained in a chamber consisting of a cut microcentrifuge tube glued to the glass mounting surface, volume $\sim 200 \mu\text{l}$) was filled with 50 μl of bacterial solution, and topped up with PBS. For dry measurements, a drop of bacterial solution was placed directly on the diamond above the NV layer, allowed to dry, rinsed with deionized water, and dried a second time. The sample was then placed in the imager with the active diamond surface facing the objective. A uniform 37-G external magnetic field was applied along a single NV axis to distinguish it from the other three NV axes. This magnetic field strength was an order of magnitude less than the coercive field typically required to flip the magnetic orientation of MTB^{3,21}, and we found that the magnetization of the MTB described here remained fixed as the external field was varied.

ODMR^{8–10} spectra were measured by imaging NV fluorescence from the whole field-of-view at different microwave frequency values. The typical total fluorescence collection time was 4 min for both wet and dry bacterial samples. For each pixel, Lorentzian fits were applied to the ODMR spectra and the magnetic field shifts along the NV axis were extracted. This procedure was repeated with the external field applied along each of the four NV axes, which in turn allowed the vector magnetic field in the NV layer to be determined for all three Cartesian directions across the field-of-view. For magnetic fields B_1 to B_4 , corresponding to measurements along axes 1 to 4, respectively, the fields in the Cartesian coordinates were calculated from

$$B_x = (3/2)^{1/2}(B_2 - B_4)/2,$$

$$B_y = (3/2)^{1/2}(B_1 - B_3)/2,$$

$$B_z = 3^{1/2}(-B_1 - B_2 - B_3 - B_4)/4$$

Bacterial viability assay. Immediately after magnetic field imaging of wet samples, the viability of the bacteria was determined in place on the diamond surface using a standard fluorescence-based live-dead assay (Molecular Probes, BacLight kit). A mixture of the fluorescent nucleic acid stains SYTO 9 (final concentration 5 μM) and propidium iodide (final concentration 30 μM) was added to the bath, and bright-field images were immediately collected to verify that the positions of the bacteria on the diamond surface were not perturbed. The sample was then incubated in the dark for 15 min, and fluorescence images were collected by exciting with a LED at 470 nm (Thorlabs). Green SYTO-9 fluorescence and red propidium iodide fluorescence were collected successively using appropriate emission filters (Thorlabs for green; Chroma for red). Custom software was used to co-register the resulting fluorescence images and perform rolling-ball background subtraction, and a peak-finding algorithm was applied to determine the positions of the bacteria. The ratio of red to green fluorescence intensity, integrated over each cell, was calculated and compared to a live/dead calibration performed previously under the same conditions (see Supplementary Information for details). MTB with a fluorescence ratio less than 0.5 were taken to be alive, while those with a fluorescence ratio greater than 1.0 were assigned as dead. Bacteria with intermediate fluorescence ratios between 0.5 and 1.0 could not be assigned to either category with high certainty based on assay calibration measurements, and were therefore labelled as indeterminate in experimental data.

Before collecting the data displayed in Fig. 2, we carried out a series of preliminary live-dead assays, including the calibrations described in Supplementary Methods. These assays revealed that, even after a full hour of exposure to $f \approx 2.88 \text{ GHz}$ microwave fields at the intensities used in our ODMR measurements, the fraction of bacteria remaining alive was essentially the same as that in unperturbed samples immediately after they were taken from culture. This suggests that any bacterial fatality during experiments was the result of residual evanescent coupling of laser light through the diamond surface. These observations were consistent with direct measurements of the bath temperature when microwave power was applied, which showed only a modest increase of 1–2 °C above room temperature.

Electron microscopy. After magnetic field measurements were completed on dried samples, imaging was performed with a field emission SEM (Zeiss Sigma). The diamond substrate and intact bacteria were carbon-coated in a thermal evaporator (Edwards Auto 306) and mounted on silicon wafers using copper tape. The bacteria were imaged without dehydration or fixation. Images of magnetic nanoparticles were obtained using backscatter mode, at $\times 30,000$ magnification and with an accelerating voltage of 8 kV. The TEM image in Fig. 1d was recorded using the procedure outlined in ref. 32.

Fitting the magnetic field of an MTB. Magnetic field patterns of the bacteria were fitted with a constrained model using SEM measurements of the relative sizes and positions of the magnetic nanoparticles, with standoff distance from the diamond and magnetic moment scaling factors left as free parameters. First, a peak-finding algorithm was applied to locate magnetic nanoparticles in the image. Magnetic nanoparticle chains were determined by assigning two adjacent magnetic nanoparticles to the same chain if their separation was less than 120 nm. For each chain, the orientation of the magnetic moment in the plane of the diamond surface was determined using a linear fit to the magnetic nanoparticle positions. Gaussian curves were fitted to the SEM images of each magnetic nanoparticle along the direction perpendicular to the axis of the chain, and the fit amplitudes were used to assign relative magnetic moment densities along the chain. Each magnetic nanoparticle in a chain was assumed to act as a point dipole with the same magnetic moment direction as its chain. (This approximation was motivated by the observation of highly aligned magnetic nanoparticle dipoles in previous work (see, for example, refs 3, 19.) In some cases, individual magnetic nanoparticle were further than 120 nm from any chains; their dipole moment was estimated to be in the same direction as that of the nearest chain.

Next, a nonlinear fit routine using the Levenberg–Marquardt algorithm was performed to match simulated magnetic field images with those measured. The simulation first calculated the three components of the magnetic field on the diamond surface using the positions, directions and relative magnetic strengths of each magnetic nanoparticle. The ODMR signal for all NV axes was then calculated for each pixel, and these signals were convolved with a point-spread function (full-width at half-maximum of 400 nm) to create simulated ODMR fluorescence data. As in the case of the measured data, images of B_x , B_y and B_z were reconstructed on a pixel-by-pixel basis from the frequency shifts for the four NV axes extracted from Lorentzian fits. The algorithm was run independently to minimize x and y position offsets of the SEM images as well as the standoff distance from the diamond surface. Generally, B_x images were used for the fitting. Finally, the overall magnetic moment was calculated on a pixel-by-pixel basis for the best-fit geometry, and the optimal value was determined by least-squares fitting to the

measured data. The best-fit magnetic moment did not depend strongly on the value of the best-fit standoff distance for typical distances of 100–200 nm, owing to convolution of the NV fluorescence signal with the ~ 400 -nm point spread function of the optical microscope. We note that this method cannot recover exact dipole orientations, particularly for isolated magnetic nanoparticles. Nevertheless, the overall magnetic moment is dominated by contributions from long chains, whose field patterns are well-described by this method.

Estimating magnetic properties directly from ODMR. In cases where magnetic nanoparticles were organized into ordered chains that were well-approximated by finite solenoids, the chain positions and magnetic moments could be determined even without comparison to SEM data. Chain locations and orientations were estimated from the measured magnetic field divergence in the diamond plane ($\partial B_x/\partial x + \partial B_y/\partial y$) by assigning chain endpoints to the local maxima and minima of the divergence. (The maximum precision of this estimate is given approximately by the diffraction-limited resolution of the ODMR measurement divided by the signal-to-noise ratio of the calculated magnetic field divergence, which is approximately 40 nm.) The chain was then approximated as a continuous line of magnetic dipoles, which can be shown to have the same field as a magnetic source and sink separated by the chain length (that is, a narrow finite solenoid). This provided a simple way to calculate B_z just below the chain. The magnetic moment could then be determined directly by spatially integrating the absolute value of B_z across the diamond surface. This integrated value is independent of standoff distance when the chain length is much larger than the standoff distance and the diameter of the field-of-view is much larger than the chain length. Moreover, it is independent of the point-spread function of the microscope objective.

31. Komeili, A., Vali, H., Beveridge, T. J. & Newman, D. K. Magnetosome vesicles are present before magnetite formation, and MamA is required for their activation. *Proc. Natl Acad. Sci. USA* **101**, 3839–3844 (2004).
32. Murat, D., Quinlan, A., Vali, H. & Komeili, A. Comprehensive genetic dissection of the magnetosome gene island reveals the step-wise assembly of a prokaryotic organelle. *Proc. Natl Acad. Sci. USA* **107**, 5593–5598 (2010).

Anomalous sulphur isotopes in plume lavas reveal deep mantle storage of Archaean crust

Rita A. Cabral¹, Matthew G. Jackson¹, Estelle F. Rose-Koga², Kenneth T. Koga², Martin J. Whitehouse^{3,4}, Michael A. Antonelli⁵, James Farquhar⁵, James M. D. Day⁶ & Erik H. Hauri⁷

Basaltic lavas erupted at some oceanic intraplate hotspot volcanoes are thought to sample ancient subducted crustal materials^{1,2}. However, the residence time of these subducted materials in the mantle is uncertain and model-dependent³, and compelling evidence for their return to the surface in regions of mantle upwelling beneath hotspots is lacking. Here we report anomalous sulphur isotope signatures indicating mass-independent fractionation (MIF) in olivine-hosted sulphides from 20-million-year-old ocean island basalts from Mangaia, Cook Islands (Polynesia), which have been suggested to sample recycled oceanic crust^{3,4}. Terrestrial MIF sulphur isotope signatures (in which the amount of fractionation does not scale in proportion with the difference in the masses of the isotopes) were generated exclusively through atmospheric photochemical reactions until about 2.45 billion years ago^{5–7}. Therefore, the discovery of MIF sulphur in these young plume lavas suggests that sulphur—probably derived from hydrothermally altered oceanic crust—was subducted into the mantle before 2.45 billion years ago and recycled into the mantle source of Mangaia lavas. These new data provide evidence for ancient materials, with negative $\Delta^{33}\text{S}$ values, in the mantle source for Mangaia lavas. Our data also complement evidence for recycling of the sulphur content of ancient sedimentary materials to the subcontinental lithospheric mantle that has been identified in diamond-hosted sulphide inclusions^{8,9}. This Archaean age for recycled oceanic crust also provides key constraints on the length of time that subducted crustal material can survive in the mantle, and on the timescales of mantle convection from subduction to upwelling beneath hotspots.

Oceanic crust and sediments are introduced to the mantle at subduction zones, but the fate of this subducted material within the mantle, as well as the antiquity of this process, is unknown. Earth's mantle is chemically and isotopically heterogeneous, and it has been suggested that some of this heterogeneity derives from geochemically diverse subducted oceanic¹ and continental² crustal material that is mixed with the ambient mantle following subduction. It has also been suggested that different types of crustal materials generate different isotopic endmembers in the mantle¹⁰—including HIMU (high $\mu = ^{238}\text{U}/^{204}\text{Pb}$), EM1 (enriched mantle I) and EM2 (enriched mantle II)—and these endmembers are sampled by mantle melts erupted at oceanic hotspot volcanoes. Owing to the loss of fluid-mobile Pb from altered basalt during subduction¹¹, oceanic crust processed in subduction zones is thought to form a HIMU reservoir in the mantle and, over time, this reservoir develops extreme radiogenic Pb-isotope compositions^{1,3,4,10}.

Basaltic lavas on the island of Mangaia exhibit the most radiogenic Pb-isotope compositions observed in ocean island basalt (OIB) globally (see, for example, refs 3 and 4) and represent the HIMU mantle endmember. Mangaia lavas have long been suggested to sample melts of recycled oceanic crust^{3,4}. However, such an origin for this signature has been questioned and alternative models that favour metasomatic

processes to generate the HIMU mantle beneath Mangaia have been suggested^{12,13}. Here we report MIF S-isotope compositions in Mangaia lavas that require the presence of recycled, ancient (>2.45 Gyr old) surface material in the HIMU mantle source for Mangaia lavas.

Fresh basaltic glass for S-isotope measurement is not available from Mangaia, where subaerial lavas are ~20 million years old¹⁴ and have suffered from extensive weathering in a tropical climate. However, magmatic olivine phenocrysts encapsulate primary magmatic sulphides and isolate them from surface weathering processes. Olivine phenocrysts were separated from three basaltic lavas collected from Mangaia. The largest inclusions were exposed for S-isotope analysis by secondary ion mass spectrometry (SIMS). Whereas sulphides <10 μm in diameter are relatively common, the largest sulphides, which permit replicate S-isotope measurements, are exceedingly rare (thousands of olivine fragments from many kilograms of rock were examined individually under a microscope and only two sulphide inclusions were large enough to permit replicate S-isotope analyses). The other sulphides were either too small for measurement by SIMS, or were sufficiently large for only a single S-isotope analysis (see Supplementary Information and Fig. 1 for sulphide descriptions). An olivine separate was prepared and analysed using chemical extraction techniques and gas-source isotope ratio mass spectrometry (IRMS) to provide an independent measurement comparison with the SIMS results.

Sulphur isotopes were measured by SIMS at the NordSIMS facility in Stockholm, Sweden, in four sulphide inclusions recovered from

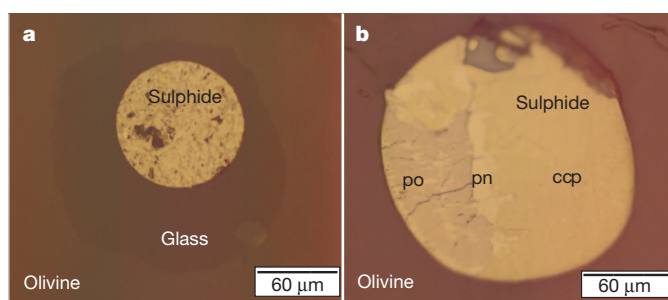


Figure 1 | Reflected-light photomicrographs of sulphide inclusions. **a**, MGA-B-47 sulphide inclusion. The sulphide was homogenized on a heating stage before exposure and analysis, and the primary magmatic sulphide mineralogy was lost during this process. **b**, MGA-B-25 sulphide inclusion. The sulphide was not homogenized, and hosts three coexisting magmatic phases: chalcopyrite (ccp), pentlandite (pn) and pyrrhotite (po). The other two sulphides examined in this study (not shown, see Supplementary Fig. 3) were separated from whole rock sample MG1001. The MG1001B-S17 sulphide inclusion contains chalcopyrite and pyrrhotite. The MG1001B-S14 sulphide inclusion, which does not have a $\Delta^{33}\text{S}$ anomaly, contains pyrrhotite, pentlandite, chalcopyrite and pyrite (a low-temperature sulphide phase consistent with a non-magmatic origin).

¹Department of Earth and Environment, Boston University, 675 Commonwealth Avenue, Boston, Massachusetts 02215, USA. ²Laboratoire Magmas et Volcans, Université Blaise Pascal, CNRS UMR6524, IRD R163, 5 rue Kessler, 63038 Clermont-Ferrand, France. ³Swedish Museum of Natural History, Box 50007, SE-104 05 Stockholm, Sweden. ⁴Department of Geological Sciences, Stockholm University, SE-106 91 Stockholm, Sweden. ⁵Department of Geology and ESSIC, University of Maryland, College Park, Maryland 20742, USA. ⁶Geosciences Research Division, Scripps Institution of Oceanography, La Jolla, California 92093-0244, USA. ⁷Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington DC 20015, USA.

three basaltic hand samples. Multiple spot analyses were made on each of the two largest sulphides, MGA-B-25 (ten spot analyses) and MGA-B-47 (nine spot analyses), and a single spot analysis was performed on each of two small sulphide inclusions from MG1001 (sulphides S14 and S17). Individual measurements of MGA-B-25 and MGA-B-47 were averaged (see discussion in Supplementary Information), and give negative $\Delta^{33}\text{S}$ anomalies (weighted averages $-0.25 \pm 0.07\text{‰}$ (2σ) and $-0.34 \pm 0.08\text{‰}$ (2σ), respectively; see Supplementary Information for discussion of uncertainty, and Supplementary Table 3) that are statistically resolvable from ambient mantle sulphur ($\Delta^{33}\text{S} = 0$). Here $\Delta^{33}\text{S} = \delta^{33}\text{S} - [(1 + \delta^{34}\text{S})^{0.515} - 1]$, $\delta^{33}\text{S}_{\text{V-CDT}} = [({}^{33}\text{S}/{}^{32}\text{S})_{\text{sample}} / ({}^{33}\text{S}/{}^{32}\text{S})_{\text{V-CDT}} - 1]$, and similarly for $\delta^{34}\text{S}$ (details of the standards are given in the Supplementary Information). The two sulphide inclusions from sample MG1001 gave overlapping $\Delta^{33}\text{S}$ values: inclusions S17 and S14 have respective $\Delta^{33}\text{S}$ values of $-0.17 \pm 0.27\text{‰}$ (2σ) and $0.03 \pm 0.28\text{‰}$ (2σ). The anomaly-free sulphide phase hosts pyrite, which is consistent with a low-temperature, non-magmatic (modern)

origin for the sulphur in this sulphide, and suggests that not all olivine-hosted sulphides in Mangaia lavas are magmatic. The $\delta^{34}\text{S}$ values of all sulphide inclusions are less than -6.1‰ (see Supplementary Information for discussion of $\delta^{34}\text{S}$ measurements), which are generally more negative than the values encountered previously in magmatic sulphide inclusions¹⁵. The possibility of dilution with normal mantle S (at $\Delta^{33}\text{S} = 0$, $\delta^{34}\text{S} = 0$) during the magmatic process means that our inclusion data probably represent mixtures, and more extreme compositions may exist in the low $\delta^{34}\text{S}$ -negative MIF source region; some evidence for this dilution comes from the apparent 'mixing line' of the inclusions, the bulk olivine and ambient mantle (Fig. 2). Alternatively, the S-isotope trend may simply reflect isotopic diversity observed in melt inclusions from Mangaia¹⁶.

Following sulphur extraction by wet chemistry, S isotopes were also measured in ~ 400 mg of bulk olivine separates from whole rock sample MGA-B-47 by gas-source IRMS at the University of Maryland. $\Delta^{33}\text{S}$ ($-0.12 \pm 0.04\text{‰}$) and $\delta^{34}\text{S}$ ($-3.28 \pm 2\text{‰}$) values are identified in the bulk olivine separate (Supplementary Table 5), but the values are smaller in magnitude than observed in the magmatic sulphides from this sample. We consider it likely that the magnitude of $\Delta^{33}\text{S}$ and $\delta^{34}\text{S}$ in the bulk olivine separates was diminished relative to the individual magmatic sulphides by incorporation of sulphur into the bulk olivine measurement, either through post-lava flow emplacement of secondary pyrite in Mangaia olivines, or through dilution of an Archaean MIF-S signature by mixing with an ambient mantle S-isotope composition.

Lead-isotope compositions of the two largest inclusions exhibiting the clearest $\Delta^{33}\text{S}$ anomalies were also measured by SIMS (Fig. 3; see Methods and Supplementary Information). Both sulphide inclusions exhibit Pb-isotope signatures indistinguishable from the whole rock Pb-isotope analyses³, confirming that the anomalous S-isotope compositions are associated with the HIMU mantle reservoir. Olivine-hosted sulphide inclusions from this locality were previously found to have endmember HIMU compositions^{16,17}.

A modern origin of the $\Delta^{33}\text{S}$ anomaly in Mangaia sulphides is improbable. Small variations in $\Delta^{33}\text{S}$ (from -0.05‰ to $+0.34\text{‰}$)

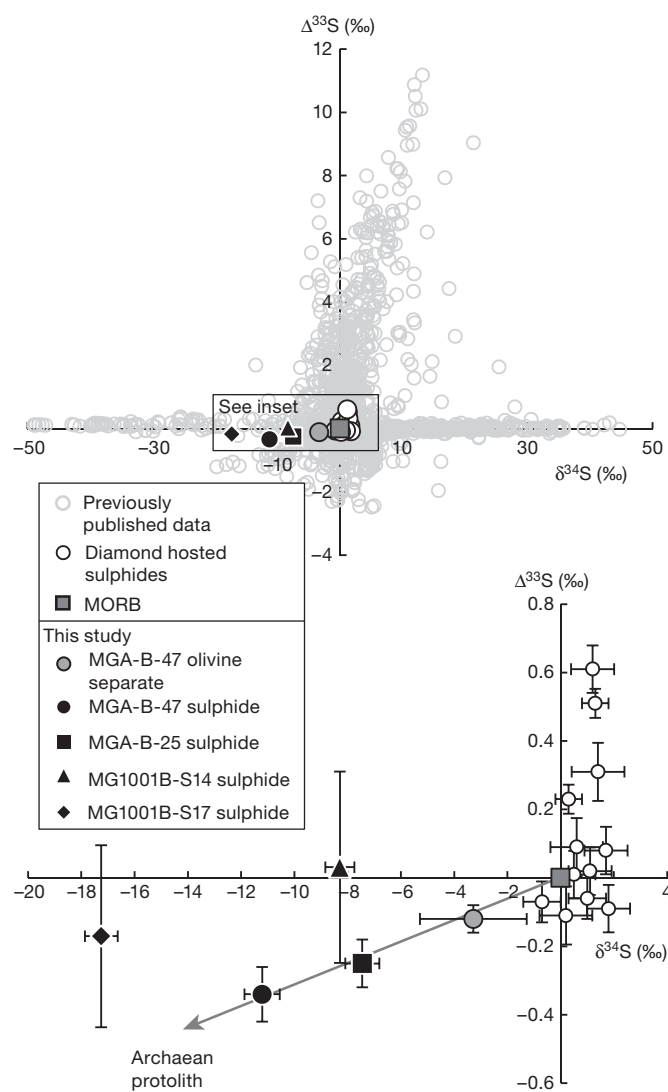


Figure 2 | $\Delta^{33}\text{S}$ versus $\delta^{34}\text{S}$ for olivine-hosted sulphide inclusions from Mangaia (this study) and diamond-hosted sulphides (from ref. 8) compared to previously published S-isotope data. Points shown are the weighted averages of the individual analyses for MGA-B-47 and MGA-B-25 ($n = 9$ for MGA-B-47; $n = 10$ for MGA-B-25) and single analyses for both MG1001 samples. Error bars are 95% confidence level for the MGA-B-25 and MGA-B-47 weighted averages and 2σ for the single analyses. The isotope composition of the bulk olivine separate is also shown (see Supplementary Information). Previously published sulphur isotope data are after figure 1 in ref. 7.

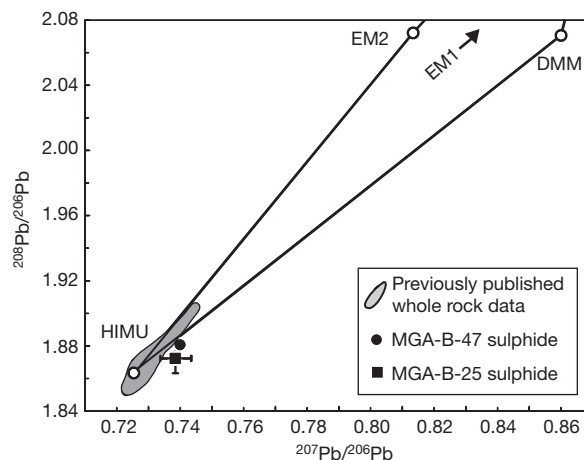


Figure 3 | The Pb-isotope composition of olivine-hosted sulphides are the same as Mangaia whole rocks. The Pb-isotope compositions of the two olivine-hosted sulphides from MGA-B-25 and MGA-B-47 were obtained by SIMS measurement (see Supplementary Information), and whole rock Pb-isotopes for these samples were characterized in ref. 3. The new sulphide data (black symbols) cluster around the HIMU mantle endmember defined by previously published whole rock Pb-isotope data from Mangaia lavas (grey field, using whole rock data from ref. 4 and references therein). The apices of the quadrilateral are defined by the isotopic endmembers found in the oceanic mantle (EM1, EM2, HIMU, DMM (depleted MORB mantle)). The average of two Pb-isotope measurements of the same inclusion are shown for MGA-B-25 (see Supplementary Table 4). Error bars reflect the 2σ standard error of the mean (MGA-B-47) or the 2σ weighted error of the mean (for the two measurements of MGA-B-25).

can be generated by biologically-controlled mass-dependent fractionation mechanisms^{18–20}, but these mechanisms tend to generate positive $\Delta^{33}\text{S}$ values when $\delta^{34}\text{S}$ values are negative, and negative $\Delta^{33}\text{S}$ when $\delta^{34}\text{S}$ values are positive, rather than the observed negative $\Delta^{33}\text{S}$ and $\delta^{34}\text{S}$ values in Mangaia sulphides. These processes are also only known to generate smaller magnitude negative $\Delta^{33}\text{S}$ anomalies than those observed in Mangaia sulphides.

Assimilation of ancient crustal materials is also an unlikely source of the $\Delta^{33}\text{S}$ anomalies. The oceanic lithosphere beneath Mangaia is too young to have formed at a time when MIF S is known to have occurred. It is also unlikely that Mangaia lavas were contaminated by stranded blocks of Archaean continental crust, as tectonic reconstructions of the Pacific plate²¹ place Mangaia far from the locus of continental rifting and from Pacific fracture zones that may have stranded ancient continental material in this oceanic basin (see, for example, refs 22 and 23).

We suggest that MIF S in Mangaia lavas comes from a mantle reservoir containing S subducted before 2.45 Gyr ago. The subducted S was preserved in the convecting mantle, and remained associated with the subducted package (subducted lithosphere \pm sediments), so that its Archaean MIF signature was not completely diluted during its >2.45 -Gyr residence in the mantle. The lower mantle may be a 'graveyard' for subducted Archaean crust with negative $\Delta^{33}\text{S}$ (ref. 8). Less vigorous convective motions in this part of the mantle may be more conducive to preserving mantle heterogeneities over long timescales. Processes associated with buoyant upwelling could have transported MIF-S-bearing material back to the surface where it melted beneath Mangaia at 20 Myr ago. Sulphides in Mangaia melts were trapped and encapsulated in growing magmatic olivine phenocrysts, thus preserving the MIF signature during magma transport and eruption.

Positive and negative $\Delta^{33}\text{S}$ have been documented previously in diamond-hosted sulphide inclusions^{8,9}, and these data complement the negative $\Delta^{33}\text{S}$ measurements reported in Mangaia. A conceptual model suggested (ref. 8) for the origin of the positive $\Delta^{33}\text{S}$ signature in the diamond-hosted sulphides sheds light on the possible origins of the negative $\Delta^{33}\text{S}$ in Mangaia. Photolysis of volcano-sourced Archaean sulphur (with initial $\Delta^{33}\text{S} = 0\text{‰}$) occurred in an oxygen-poor (and therefore ozone-poor) atmosphere relatively transparent to solar ultraviolet radiation. Photochemical fractionation acting on atmospheric sulphur species generated geochemical reservoirs with complementary positive $\Delta^{33}\text{S}$ (reduced and elemental sulphur species) and negative $\Delta^{33}\text{S}$ (oxidized species such as sulphate). Elemental sulphur with positive $\Delta^{33}\text{S}$ in the atmosphere was deposited in surface reservoirs and converted to sulphide. The positive $\Delta^{33}\text{S}$ of the sulphur identified in the diamond-hosted sulphide inclusions corresponds to that found in Archaean sedimentary sulphides, suggesting a sedimentary origin for the positive $\Delta^{33}\text{S}$ in the diamond-hosted sulphides⁸. The Archaean sulphur-bearing sediment was subducted into the mantle source of the Orapa diamonds, encapsulated in diamond, and preserved until transport to the surface in a kimberlite eruption.

The data reported here include the first observation of non-zero $\Delta^{33}\text{S}$ in OIB. We suggest that the negative $\Delta^{33}\text{S}$ identified in Mangaia sulphides originates from the Archaean oceanic sulphate pool that is complementary to the sedimentary pyrite, and that this MIF S-isotope signature was incorporated into oceanic crust following bisulphide formation in Archaean hydrothermal systems (Supplementary Fig. 1). Archaean rocks with a clear oceanic association tend to exhibit negative $\Delta^{33}\text{S}$ signatures. Indeed, hydrothermally-influenced Archaean deposits tend to exhibit negative $\Delta^{33}\text{S}$ (refs 24–28). Subduction of hydrothermally-altered Archaean basalt into the mantle can produce a negative $\Delta^{33}\text{S}$ reservoir of subducted oceanic lithosphere in the deep mantle that may also help to explain the apparent bias to positive $\Delta^{33}\text{S}$ values seen in compilations of published analyses^{6,7}. Therefore, the negative $\Delta^{33}\text{S}$ of sulphides analysed here point to an ancient crustal source with oceanic affinities for Mangaia sulphur. An origin associated with Archaean crustal material is also consistent with

other geochemical characteristics of the Mangaia lavas, such as the radiogenic Pb-isotope compositions^{1,3,4,10}.

The S-isotope compositions from Mangaia require a protolith with the unusual combination of negative $\Delta^{33}\text{S}$ and negative $\delta^{34}\text{S}$ values. Whereas the negative $\Delta^{33}\text{S}$ can only have been generated in the Archaean atmosphere, the origin of the negative $\delta^{34}\text{S}$ signature in Mangaia sulphides is less well-constrained. Archaean volcanogenic massive sulphide (VMS) deposits hosted in komatiites represent a possible candidate for the recycled protolith melted beneath Mangaia, as such deposits can have both the negative $\Delta^{33}\text{S}$ and negative $\delta^{34}\text{S}$ (ref. 24) that approach those identified in Mangaia sulphides. Komatiites may have been commonly erupted on the seafloor during the Archaean²⁹, where they could have incorporated negative $\Delta^{33}\text{S}$ values by seawater sulphate reduction at hydrothermal settings, as evidenced by some Archaean VMS deposits²⁴. Therefore, one possible model for the S-isotope composition of Mangaia sulphides is that it originates in hydrothermally-modified mafic sources, similar to the komatiite-hosted VMS deposits²⁴, that were subsequently subducted into the mantle during the Archaean.

Whereas VMS deposits trend in the direction of negative $\Delta^{33}\text{S}$ and negative $\delta^{34}\text{S}$ identified in Mangaia sulphides, the available S-isotope data on VMS deposits do not extend to the low $\delta^{34}\text{S}$ values we observe in Mangaia. Therefore, we cannot exclude alternative mechanisms that might have generated the combination of negative $\Delta^{33}\text{S}$ and negative $\delta^{34}\text{S}$ in the Archaean. The combination of negative $\Delta^{33}\text{S}$ and $\delta^{34}\text{S}$ (down to -0.71‰ and -8.3‰ , respectively) was identified in Archaean sulphides from the Gamohaan formation, South Africa³⁰, where negative $\delta^{34}\text{S}$ was attributed to bacterial reduction of sulphate (a process demonstrated to generate extreme negative $\delta^{34}\text{S}$ signatures³¹) while preserving its negative $\Delta^{33}\text{S}$ anomaly inherited from photolytic reactions in the atmosphere. High degrees of melt degassing under reducing conditions might also generate highly negative $\delta^{34}\text{S}$ values, but this would require the melt to have initially negative $\Delta^{33}\text{S}$, inherited from the Archaean atmosphere, so that the final degassed product has the combination of S-isotope compositions observed in Mangaia inclusions. Although the exact mechanism for generating negative $\delta^{34}\text{S}$ is unknown, the key feature in the Mangaia data set is that the sulphide inclusions and bulk olivine separate have $\Delta^{33}\text{S}$ anomalies that could only have been generated in the Archaean atmosphere.

The identification of MIF S in Mangaia lavas places two critical constraints on the origin of the HIMU mantle. First, several recent models invoke metasomatic processes occurring within the mantle to generate the HIMU reservoir (see, for example, refs 12 and 13), but such models do not explicitly invoke materials recycled from surface reservoirs and therefore cannot explain MIF S in HIMU lavas. The discovery of MIF S requires subduction of surface materials into the mantle to generate the HIMU reservoir sampled by Mangaia lavas. Second, the $\Delta^{33}\text{S}$ anomaly associated with Mangaia lavas places a lower limit on the formation age of the HIMU mantle domain, namely, 2.45 Gyr ago (Supplementary Fig. 2). The Archaean age for HIMU formation indicated by S isotopes conflicts with earlier estimates that are based on two-stage Pb-isotope model ages of ~ 1.8 Gyr ago (ref. 10), and this discrepancy may imply a more complicated history for Pb isotopes (that is, more than two stages of Pb differentiation are possible) than generally assumed for the generation of the HIMU mantle. The >2.45 -Gyr age constraint from S isotopes indicates that mantle heterogeneities generated by subduction of surface materials into the mantle can be preserved over long timescales—from the Archaean to present—in the convecting mantle.

The new $\Delta^{33}\text{S}$ measurements confirm inferences about the cycling of sulphur between the major reservoirs from the Archaean to the Phanerozoic, extending from the atmosphere and oceans to the crust and mantle, and ultimately through a return cycle to the surface that, here, is completed in Mangaia lavas. It remains to be seen whether lavas erupted at other HIMU hotspots and hotspots sampling different

compositional mantle endmembers (for example, EM1 and EM2) will exhibit evidence for recycling of Archaean protoliths.

METHODS SUMMARY

Three basaltic rock samples from Mangaia were crushed, sieved and picked for olivines hosting melt inclusions. Two of the four sulphides analysed here were homogenized using two different techniques (see Methods), and the other two sulphides were not homogenized. Following exposure of the sulphides, the samples were pressed in indium, dried in a furnace, cleaned and then gold coated for SIMS analyses.

In situ Pb- and S-isotope measurements were made using a CAMECA IMS 1280 SIMS instrument at the Swedish Museum of Natural History, Stockholm (NordSIMS facility). All four isotopes of Pb were measured at a mass resolution of 4,860 ($M/\Delta M$) by static four-electron multiplier configuration using a $^{16}\text{O}_2^-$ primary ion beam with 23 kV incident energy. Corrections for instrumental mass fractionation were made using natural basaltic glass standards.

The three most abundant S-isotopes (^{32}S , ^{33}S and ^{34}S) were measured at a mass resolution of 4,860 ($M/\Delta M$; sufficient to resolve $^{33}\text{S}^-$ from $^{32}\text{S}^1\text{H}^-$) by static multicollection on Faraday detectors using a $^{133}\text{Cs}^+$ primary beam with an incident energy of 20 kV. Corrections for instrumental mass fractionation were made using natural sulphide standards. Full procedures for Pb- and S-isotope analysis at NordSIMS are outlined in the Methods.

Sulphur isotope measurements were made on sulphide inclusions in acid-washed olivine separates from sample MGA-B-47. Sulphur was extracted using Cr reduction techniques and converted to silver sulphide that was analysed as SF_6 by gas-source IRMS at the University of Maryland using techniques described in the Methods.

Major element compositions of sulphides, host-olivine and glass were determined using a CAMECA SX100 electron microprobe at the Laboratoire Magmas et Volcans, Clermont-Ferrand, France.

Full Methods and any associated references are available in the online version of the paper.

Received 16 October 2012; accepted 14 February 2013.

- Hofmann, A. W. & White, W. M. Mantle plumes from ancient oceanic crust. *Earth Planet. Sci. Lett.* **57**, 421–436 (1982).
- White, W. & Hofmann, A. Sr and Nd isotope geochemistry of oceanic basalts and mantle evolution. *Nature* **296**, 821–825 (1982).
- Hauri, E. & Hart, S. R. Re-Os isotope systematics of HIMU and EMII oceanic island basalts from the south Pacific Ocean. *Earth Planet. Sci. Lett.* **114**, 353–371 (1993).
- Hanyu, T. *et al.* Geochemical characteristics and origin of the HIMU reservoir: a possible mantle plume source in the lower mantle. *Geochim. Geophys. Geosyst.* **12**, Q0AC09, <http://dx.doi.org/10.1029/2010GC003252> (2011).
- Farquhar, J., Bao, H. & Thiemens, M. Atmospheric influence of Earth's earliest sulfur cycle. *Science* **289**, 756–758 (2000).
- Farquhar, J., Zerkle, A. L. & Bekker, A. Geological constraints on the origin of oxygenic photosynthesis. *Photosynth. Res.* **107**, 11–36 (2011).
- Johnston, D. T. Multiple sulfur isotopes and the evolution of Earth's surface sulfur cycle. *Earth Sci. Rev.* **106**, 161–183 (2011).
- Farquhar, J., Wing, B., McKeegan, K. & Harris, J. Mass-independent sulfur of inclusions in diamond and sulfur recycling on early Earth. *Science* **298**, 2369–2372 (2002).
- Thomassot, E. *et al.* Metasomatic diamond growth: a multi-isotope study (^{13}C , ^{15}N , ^{33}S , ^{34}S) of sulphide inclusions and their host diamonds from Jwaneng (Botswana). *Earth Planet. Sci. Lett.* **282**, 79–90 (2009).
- Zindler, A. & Hart, S. Chemical geodynamics. *Annu. Rev. Earth Planet. Sci.* **14**, 493–571 (1986).
- Kelley, K. A., Plank, T., Farr, L., Ludden, J. & Staudigel, H. Subduction cycling of U, Th, and Pb. *Earth Planet. Sci. Lett.* **234**, 369–383 (2005).
- Niu, Y. & O'Hara, M. J. Origin of ocean island basalts: a new perspective from petrology, geochemistry, and mineral physics considerations. *J. Geophys. Res.* **108**, 2209–2228 (2003).
- Pilet, S., Baker, M. B. & Stolper, E. M. Metasomatized lithosphere and the origin of alkaline lavas. *Science* **320**, 916–919 (2008).
- Turner, D. & Jarrard, R. K-Ar dating of the Cook-Austral island chain: a test of the hot-spot hypothesis. *J. Volcanol. Geotherm. Res.* **12**, 187–220 (1982).
- Chaussidon, M., Albarède, F. & Sheppard, S. M. F. Sulphur isotope variations in the mantle from ion microprobe analyses of micro-sulphide inclusions. *Earth Planet. Sci. Lett.* **92**, 144–156 (1989).
- Saal, A. E., Hart, S. R., Shimizu, N., Hauri, E. H. & Layne, G. D. Pb isotopic variability in melt inclusions from oceanic island basalts, Polynesia. *Science* **282**, 1481–1484 (1998).
- Yurimoto, H. *et al.* Lead isotopic compositions in olivine-hosted melt inclusions from HIMU basalts and possible link to sulfide components. *Phys. Earth Planet. Inter.* **146**, 231–242 (2004).
- Ono, S., Wing, B., Johnston, D., Farquhar, J. & Rumble, D. Mass-dependent fractionation of quadruple stable sulfur isotope system as a new tracer of sulfur biogeochemical cycles. *Geochim. Cosmochim. Acta* **70**, 2238–2252 (2006).
- Shen, Y. *et al.* Multiple S-isotopic evidence for episodic shoaling of anoxic water during Late Permian mass extinction. *Nature Commun.* **2**, 210 (2011).
- Rouxel, O., Ono, S., Alt, J., Rumble, D. & Ludden, J. Sulfur isotope evidence for microbial sulfate reduction in altered oceanic basalts at ODP Site 801. *Earth Planet. Sci. Lett.* **268**, 110–123 (2008).
- Taylor, B. The single largest oceanic plateau: Ontong Java-Manihiki-Hikurangi. *Earth Planet. Sci. Lett.* **241**, 372–380 (2006).
- Frey, F. *et al.* Origin and evolution of a submarine large igneous province: the Kerguelen Plateau and Broken Ridge, southern Indian Ocean. *Earth Planet. Sci. Lett.* **176**, 73–89 (2000).
- Kamenetsky, V., Maas, R. & Sushchevskaya, N. Remnants of Gondwanan continental lithosphere in oceanic upper mantle: evidence from the South Atlantic Ridge. *Geology* **29**, 243–246 (2001).
- Bekker, A. *et al.* Atmospheric sulfur in Archean komatiite-hosted nickel deposits. *Science* **326**, 1086–1089 (2009).
- Farquhar, J. & Wing, B. in *Mineral Deposits and Earth Evolution* (eds McDonald, I., Boyce, A. J., Butler, I. B., Herrington, R. J. & Polya, D. A.) 167–177 (Spec. Publ. 248, Geol. Soc. Lond., 2005).
- Ueno, Y., Ono, S., Rumble, D. & Maruyama, S. Quadruple sulfur isotope analysis of ca. 3.5 Ga Dresser Formation: new evidence for microbial sulfate reduction in the early Archean. *Geochim. Cosmochim. Acta* **72**, 5675–5691 (2008).
- Bao, H., Rumble, D. & Lowe, D. R. The five stable isotope compositions of Fig Tree barites: implications on sulfur cycle in ca. 3.2 Ga oceans. *Geochim. Cosmochim. Acta* **71**, 4868–4879 (2007).
- Jamieson, J., Wing, B., Hannington, M. & Farquhar, J. Evaluating isotopic equilibrium among sulfide mineral pairs in Archean ore deposits: case study from the Kidd Creek VMS deposit, Ontario, Canada. *Econ. Geol.* **101**, 1055–1061 (2006).
- Grove, T. L. & Parman, S. W. Thermal evolution of the Earth as recorded by komatiites. *Earth Planet. Sci. Lett.* **219**, 173–187 (2004).
- Kamber, B. S. & Whitehouse, M. J. Micro-scale sulphur isotope evidence for sulphur cycling in the late Archean shallow ocean. *Geobiology* **5**, 5–17 (2007).
- McLoughlin, N., Grosch, E. G., Kilburn, M. R. & Wacey, D. Sulfur isotope evidence for a Paleoproterozoic subsurface biosphere, Barberton, South Africa. *Geology* **40**, 1031–1034 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements M.G.J. acknowledges Boston University start-up funds and NSF grant EAR-1145202 that supported this work. E.F.R.-K. and K.T.K. acknowledge support from EU SYNTHESIS and French ANR SlabFlux, this is Laboratory of Excellence ClerVolc contribution number 54. The NordSIMS facility is financed and operated under a joint Nordic contract; this is NordSIMS contribution number 337. We thank B. White for his review of the manuscript, and P. Cartigny and J. Labidi for discussion. We thank D.T. Johnston, D. Papineau, O. J. Rouxel and S. Ono for advice in compiling the global S-isotope database. We also thank N. Shimizu, B. D. Monteleone, E. A. Price, and P. Schiano for assistance with sample preparation.

Author Contributions R.A.C. wrote the paper and prepared the figures and tables. M.G.J. conceived the project. R.A.C., E.F.R.-K., K.T.K. and M.G.J. performed sample preparation. M.J.W. performed the *in situ* SIMS analyses. J.F. and M.A.A. performed the S-isotope analyses on bulk olivine separates. J.M.D.D. and E.H.H. aided in the field. All authors participated in the discussion and interpretation of results, and preparation of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.A.C. (racabral@bu.edu).

METHODS

Sample preparation and selection. Three basaltic samples (MG1001, MGA-B-25 and MGA-B-47) were crushed and sieved. Olivines hosting melt inclusions were picked using a binocular microscope. One olivine-hosted melt inclusion (sample MGA-B-47) was homogenized using a Vernadsky-type heating stage while monitoring the homogenization temperature of the inclusion. Batches of 50–150 mg of olivine crystals from sample MG-1001 were homogenized using a gas-mixing furnace at 1,280 °C for 20 min at a near iron–wüstite buffer condition. Duration and oxygen fugacity of this batch homogenization technique closely mimicked the homogenization procedure by the heating stage. Two sulphide inclusions were homogenized, but the other two (MGA-B-25 and MG1001-S14) were not. Following exposure of the sulphides by polishing, sulphides were pressed into an indium mount, re-polished, cleaned with deionized water, and dried at 100 °C for 24 h before applying a gold coat.

In situ lead isotope measurements. *In situ* Pb-isotope measurements of the sulphides used methods described elsewhere³². A $-13\text{ kV }^{16}\text{O}_2^-$ primary ion beam illuminated a 200 μm mass aperture to produce a $\sim 7\text{--}8\text{ nA}$, 20 μm , slightly elliptical, flat bottomed crater. Target areas were subjected to a 180 s pre-sputter with a $25 \times 25\text{ }\mu\text{m}$ raster to remove the Au coating and clean the surface of extraneous Pb. The unrastered 10 kV secondary ion beam was centred in a 4,000 μm field aperture (field of view on the sample approximately $25 \times 25\text{ }\mu\text{m}$ at 160 \times transfer magnification) by scanning the transfer deflectors, and the beam was maximized to the peak of the energy distribution (45 eV window) by scanning the sample voltage. The magnetic field was locked at high precision using an NMR field sensor throughout the analytical session. The mass spectrometer used an entrance slit width of 60 μm and a common exit slit width of 250 μm on four ion counting secondary electron multipliers (EMs), corresponding to a mass resolution ($M/\Delta M$) of 4860, sufficient to resolve Pb from molecular interferences in sulphides and glasses. The detectors were positioned for simultaneous detection of ^{204}Pb , ^{206}Pb , ^{207}Pb and ^{208}Pb . Measurements consisted of 40–120 cycles of 20 s integration. An electronically gated 60 ns deadtime correction was used. Typical background levels on the EMs were $<0.02\text{ c.p.s.}$, which was negligible at the level of Pb signal measured. Each set of 40 cycles took $\sim 20\text{ min}$. See Supplementary Information for discussion of internal and external precision for *in situ* analyses of Pb isotopes.

In situ sulphur-isotope measurements. Multiple S-isotope measurements followed the analytical protocol described by ref. 33. A -10 kV primary beam of $^{133}\text{Cs}^+$ was critically focused onto the sample, yielding a 2 nA primary beam and spot with a diameter of $\sim 5\text{ }\mu\text{m}$ which was rastered over $5 \times 5\text{ }\mu\text{m}$ during data acquisition to homogenize the sampling. Target areas were subjected to a 70 s pre-sputter with a $20 \times 20\text{ }\mu\text{m}$ raster to remove the Au coating. The $5 \times 5\text{ }\mu\text{m}$ rastered 10 kV secondary ion beam was centred in a 2,500 μm field aperture (field of view on the sample of $\sim 25 \times 25\text{ }\mu\text{m}$ at 100 \times transfer magnification) by automated scanning of the transfer deflectors. Sample charging was minimized by use of a low-energy normal-incidence electron gun and no energy adjustments were necessary. The magnetic field was locked at high precision using an NMR field sensor for the entire analytical session. The mass spectrometer used an entrance slit width of 90 μm and a common exit slit width of 250 μm on the three Faraday detectors used to measure ^{32}S , ^{33}S and ^{34}S , corresponding to a mass resolution ($M/\Delta M$) of 4,860. Faraday amplifiers were housed in an evacuated, thermally stabilized

chamber and used a $10^{10}\text{ }\Omega$ input resistor on the ^{32}S channel and $10^{11}\text{ }\Omega$ on the other channels. Typical secondary ion signals of 10^9 c.p.s. on ^{32}S were obtained and each analysis consisted of 64 s of data integration.

The S-isotope data were obtained in two analytical sessions. Analyses of the unknown sulphides were bracketed by measurements of two non-MIF pyrite standards, Ruttan and Balmat³⁴ and a MIF pyrite from the Isua Greenstone Belt³³. Ruttan alone was used for calculation of instrumental mass fractionation while Ruttan, Balmat and the Isua pyrite were used to constrain the mass dependent fractionation line for each session. In the second session, an in-house, strongly negative $\delta^{34}\text{S}$ pyrite concretion (Gabon) was used to further verify the mass dependent fractionation line, but was not used to calculate instrumental mass fractionation or to constrain the mass dependent fractionation line. See Supplementary Information for external precision of S-isotope measurements.

Sulphur-isotope measurements on bulk olivines. Sulphur isotope measurement of sulphide inclusions in $\sim 400\text{ mg}$ of olivine separates from sample MGA-B-47 were extracted using chemical techniques and measured by gas source isotope ratio mass spectrometry at the University of Maryland. Acid washed (HF and HCl) olivine separates were crushed in an agate mortar under ethanol and transferred to an apparatus like that described in ref. 35 where they were reacted with a hot acidic Cr(II) solution. Sulphide released in this process was captured as silver sulphide, which was washed, dried and wrapped in clean Al foil. The foil with silver sulphide was placed in a Ni tube where it was reacted with fluorine gas, overnight at 250 °C. Product SF_6 was purified using cryogenic and chromatographic techniques, and frozen into a micro inlet on a ThermoFinnigan MAT 253, which was used for determination of isotope ratios. The sample size was small (100 μg) and the fluorination yield was low (50%), suggesting either adsorbed water on the Al foil or possible contaminants (for example, Al oxide coatings and oils that may not have been completely cleaned off of the foil) in the silver sulphide. Uncertainties for $\Delta^{33}\text{S}$ are inferred from the mass spectrometry analyses, which yielded $\pm 0.04\text{‰}$ (2σ). Uncertainties for $\delta^{34}\text{S}$ are estimated to be larger ($\pm 1\text{‰}$, 2σ) due to contributions of mass dependent fractionation during the chemical preparation (extraction, conversion and purification) of sulphur or mass spectrometric analysis. The long-term reproducibility on fluorination is 0.016 ($2\sigma\text{ s.d.}$) for $\Delta^{33}\text{S}$ and 0.30 ($2\sigma\text{ s.d.}$) for $\delta^{34}\text{S}$. Short-term reproducibility can be better for $\delta^{34}\text{S}$ and in rare cases for $\Delta^{33}\text{S}$.

Major element measurements. Electron microprobe analysis for the determination of major element concentrations was completed at the Laboratoire Magmas et Volcans, Clermont-Ferrand, France, on a Cameca SX 100 using standard procedures.

32. Rose-Koga, E. F. *et al.* Mantle source heterogeneity for South Tyrrhenian magmas revealed by Pb isotopes and halogen contents of olivine-hosted melt inclusions. *Chem. Geol.* **334**, 266–279 (2012).
33. Whitehouse, M. J. Multiple sulfur isotope determination by SIMS: evaluation of reference sulfides for $\Delta^{33}\text{S}$ with observations and a case study on the determination of $\Delta^{36}\text{S}$. *Geostand. Geoanal. Res.* <http://dx.doi.org/10.1111/j.1751-908X.2012.00188.x> (published online 7 January 2013).
34. Crowe, D. E. & Vaughan, R. G. Characterization and use of isotopically homogeneous standards for *in situ* laser microprobe analysis of $^{34}\text{S}/^{32}\text{S}$ ratios. *Am. Mineral.* **81**, 187–193 (1996).
35. Forrest, J. & Newman, L. Ag-110 microgram sulphate analysis for short time resolution of ambient levels of sulphur aerosol. *Anal. Chem.* **49**, 1579–1584 (1977).

A systematic genome-wide analysis of zebrafish protein-coding gene function

Ross N. W. Kettleborough^{1*}, Elisabeth M. Busch-Nentwich^{1*}, Steven A. Harvey^{1*}, Christopher M. Dooley¹, Ewart de Bruijn², Freek van Eeden³, Ian Sealy¹, Richard J. White¹, Colin Herd¹, Isaac J. Nijman², Fruzsina Fényes¹, Selina Mehroke¹, Catherine Scchill¹, Richard Gibbons¹, Neha Wali¹, Samantha Carruthers¹, Amanda Hall¹, Jennifer Yen¹, Edwin Cuppen² & Derek L. Stemple¹

Since the publication of the human reference genome, the identities of specific genes associated with human diseases are being discovered at a rapid rate. A central problem is that the biological activity of these genes is often unclear. Detailed investigations in model vertebrate organisms, typically mice, have been essential for understanding the activities of many orthologues of these disease-associated genes. Although gene-targeting approaches^{1–3} and phenotype analysis have led to a detailed understanding of nearly 6,000 protein-coding genes^{3,4}, this number falls considerably short of the more than 22,000 mouse protein-coding genes⁵. Similarly, in zebrafish genetics, one-by-one gene studies using positional cloning⁶, insertional mutagenesis^{7–9}, antisense morpholino oligonucleotides¹⁰, targeted re-sequencing^{11–13}, and zinc finger and TAL endonucleases^{14–17} have made substantial contributions to our understanding of the biological activity of vertebrate genes, but again the number of genes studied falls well short of the more than 26,000 zebrafish protein-coding genes¹⁸. Importantly, for both mice and zebrafish, none of these strategies are particularly suited to the rapid generation of knockouts in thousands of genes and the assessment of their biological activity. Here we describe an active project that aims to identify and phenotype the disruptive mutations in every zebrafish protein-coding gene, using a well-annotated zebrafish reference genome sequence^{18,19}, high-throughput sequencing and efficient chemical mutagenesis. So far we have identified potentially disruptive mutations in more than 38% of all known zebrafish protein-coding genes. We have developed a multi-allelic phenotyping scheme to efficiently assess the effects of each allele during embryogenesis and have analysed the phenotypic consequences of over 1,000 alleles. All mutant alleles and data are available to the community and our phenotyping scheme is adaptable to phenotypic analysis beyond embryogenesis.

Over the past 9 years we have aimed to establish methods for the systematic identification of disruptive mutations in zebrafish. Given the lack of complete annotation at early stages of the project we originally used a reverse genetic approach known as TILLING to identify mutations in specific genes by sequencing polymerase chain reaction (PCR)-amplified exons from thousands of *N*-ethyl-*N*-nitrosourea (ENU) mutagenized individuals^{11–13,20}. With the advent of high-throughput sequencing methods we were able to substantially increase the throughput of this approach but never reached genome-wide coverage of exons¹¹.

After the release of the Zv8 and Zv9 assemblies of the zebrafish genome and their protein annotations, we were able to design reagents to extract the annotated exons from zebrafish genomic DNA, enriching for approximately 60 megabase pairs (Mbp) of exome sequence and covering all 26,206 protein-coding genes^{18,19} (Fig. 1). We increased sequencing throughput by pre-capture pooling up to eight barcoded F₁ genomic libraries and combining the exon-enriched DNA into a

single high-throughput sequencing sample, while retaining adequate sequencing coverage and depth to identify heterozygous induced mutations (Supplementary Table 1). By sequencing the exon-enriched DNA from mutagenized F₁ individuals, we are able to identify ENU-induced mutations using a modified version of the 1000 Genomes

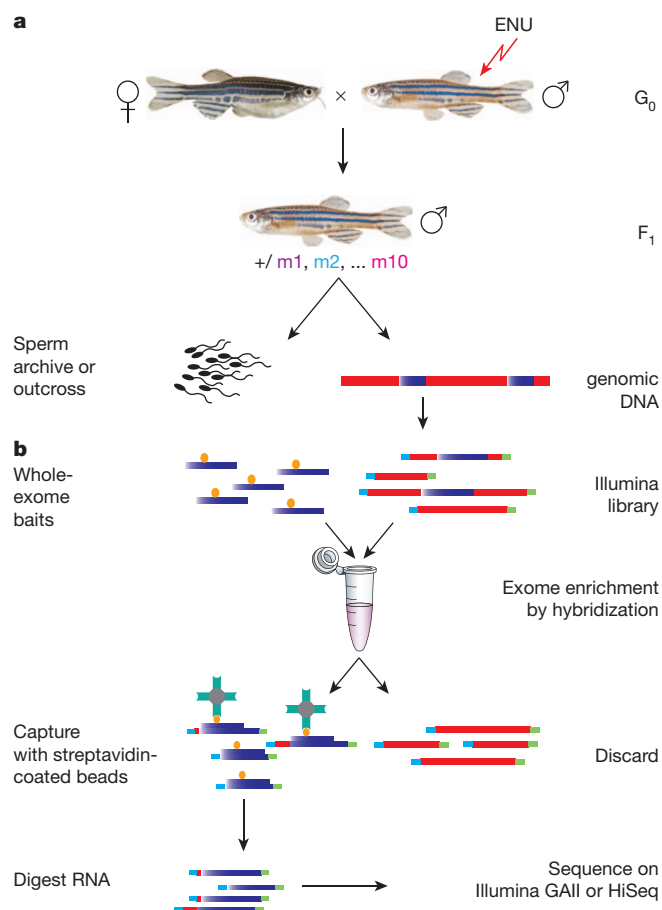


Figure 1 | Exome sequencing. **a**, ENU-mutagenized G₀ males are outcrossed to create a population of F₁ individuals heterozygous for induced mutations. Genomic DNA is taken from F₁ individuals, and the F₁ individuals are either outcrossed or cryopreserved as sperm samples. **b**, F₁ genomic DNA is then subjected to exome sequencing. Illumina libraries are made and hybridized to the 120-base biotinylated RNA oligonucleotide whole-exome baits. Streptavidin-coated magnetic beads capture genomic DNA hybridized to the RNA baits, and all other DNA is discarded. Exome-enriched DNA fragments are sequenced. Blue, exonic genomic DNA; red, non-coding genomic DNA. The designations +/m1, m2, ... m10 represent induced heterozygous mutations.

¹Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ²Hubrecht Institute, KNAW and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands. ³MRC-CDBG/Department of Biomedical Science, The University of Sheffield, Western Bank, Sheffield S10 2TN, UK.

*These authors contributed equally to this work.

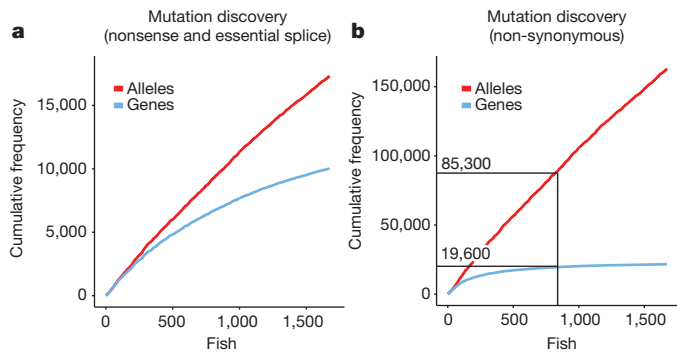


Figure 2 | Mutation detection. **a**, The cumulative detection of nonsense and essential splice alleles. As each mutagenized library displayed different rates of mutagenesis, the order that exomes were sequenced was randomized. **b**, The detection of non-synonymous mutations. Sequencing 808 exomes resulted in the identification of 85,338 non-synonymous alleles in 19,655 genes corresponding to 75% of all protein-coding genes.

project variant-calling pipeline^{21,22}. For each F_1 we predict the protein-coding consequences of the induced mutations, and for each nonsense and essential splice-site mutation we generate a single nucleotide polymorphism (SNP) genotyping assay¹¹ to facilitate the identification of each mutation in subsequent generations. We are able to confirm 95% of candidate mutations in subsequent generations.

We analysed the exome sequences of 1,673 mutagenized F_1 individuals and identified 12,002 induced nonsense and 5,337 induced essential splice mutations in 10,043 genes. For 4,105 genes we identified two or more disruptive alleles (Supplementary Table 2). With this set of data we can make predictions about the number of F_1 individuals we will need to analyse to obtain disruptive mutations in

each protein-coding gene. The detection of nonsense and essential splice mutations in new genes gradually decreased over the 1,673 sequenced individuals (Fig. 2a), such that the ratio of mutations in new genes to alleles was 1 for the first 10 sequenced exomes but 0.37 for the last 10 exomes. On average, each individual contained 125 nonsense and 168 essential splice mutations, which were common to the strains used, and among the induced mutations were 7 nonsense, 3 essential splice and 90 non-synonymous mutations (Supplementary Table 3). As the number of induced non-synonymous mutations within each individual was approximately 10 times the sum of nonsense and essential splice mutations (Fig. 2b), the rate of detecting non-synonymous mutations in new genes decreased more rapidly over the 1,673 exomes, with the ratio of non-synonymous mutations in new genes to alleles being 0.024 for the last 10 exomes. Sequencing 808 individuals resulted in the identification of 85,338 non-synonymous alleles, corresponding to mutations in 75% of all known protein-coding genes. We predict that at least 1 disruptive mutation will be identified in 75% of protein-coding genes by sequencing approximately 8,000 F_1 individuals.

To draw the most value from this resource, it is important to identify any phenotypic consequence of homozygous mutations. Thus, we have established a high-throughput, systematic phenotypic analysis of alleles to assess the developmental consequences of any given mutation. We have focused our efforts on induced nonsense and essential splice mutations. In a two-step, multi-allelic, phenotyping approach, we first identify those mutations that do not cause a phenotype within the first 5 days post fertilization (d.p.f.) in F_3 embryos (Fig. 3) collected from crosses of up to 12 pairs of F_2 individuals (Fig. 3b). We genotype phenotypically normal F_3 embryos at 5 d.p.f. for all mutations identified as heterozygous in both F_2 parents (Fig. 3c). Homozygous mutations present in the expected Mendelian ratios among F_3 embryos

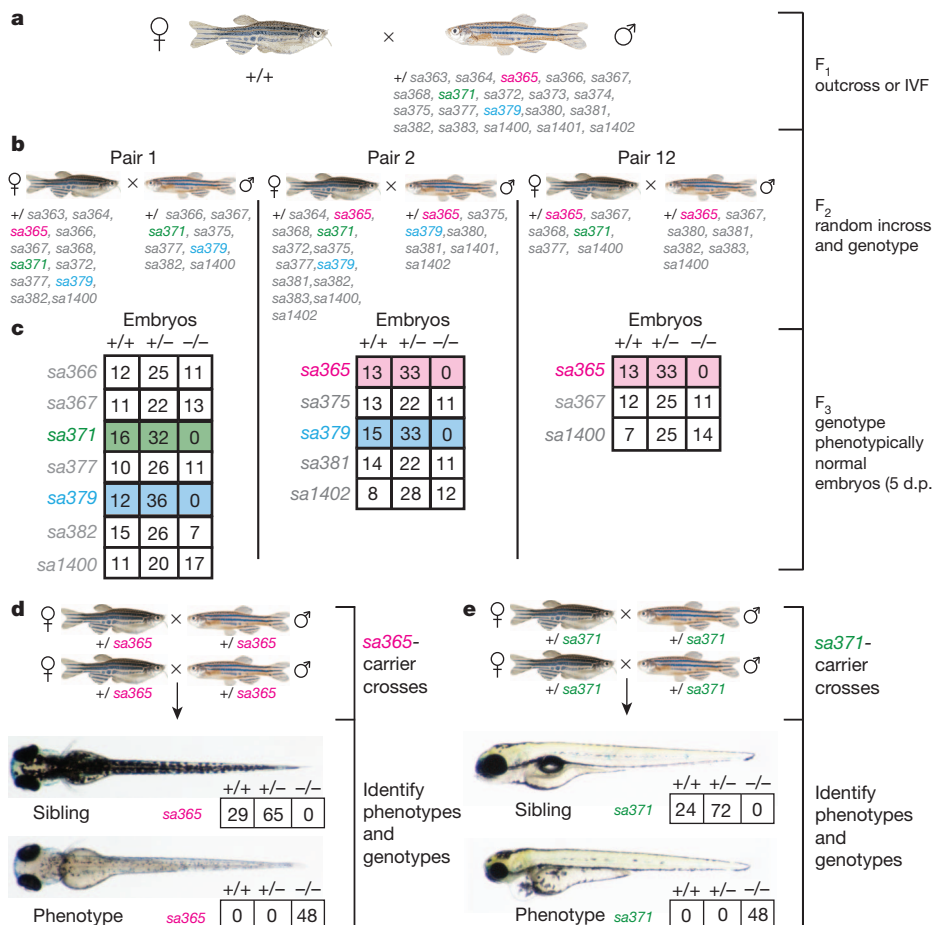


Figure 3 | Phenotypic analysis of alleles. **a**, F_1 individuals were outcrossed to produce an F_2 family. The induced disruptive alleles for one family are shown. **b**, F_2 individuals were incrossed and genotyped. **c**, First round, embryos with wild-type phenotypes were collected from each clutch at 5 d.p.f. and genotyped for the mutations heterozygous in both parents. The number of homozygous mutant F_3 embryos was assessed using a chi-squared test (P values of less than 0.05 were considered statistically significant). Mutations homozygous in less than 25% of embryos were suspected to cause a phenotype. In the wild-type phenotype set there were no homozygous embryos for the alleles *slc22a7b*^{sa365}, *mphosph10*^{sa371} and *lamc1*^{sa379}. **d**, **e**, Second round, phenotypes present within each incross were genotyped for putative phenotypic mutations. **d**, *slc22a7b*^{sa365} shows pigment phenotype. **e**, *mphosph10*^{sa371} shows small head and pericardiac oedema phenotype. *lamc1*^{sa379} mutants are shown in Fig. 4l, n.

are documented as not causing a phenotype at 5 d.p.f. Homozygous mutations present in less than 25% of phenotypically normal embryos are suspected to cause a phenotype (Fig. 3c).

In the second step of analysis, we test for correlations between the predicted disruptive mutations and morphological phenotypes (Fig. 3d, e). We re-cross the F_2 adults that are heterozygous for the suspected causal mutation and examine the F_3 embryos for all morphological and behavioural phenotypes during the first 5 d.p.f. All phenotypes are genotyped for the given mutation and if, for a given phenotype, over 90% of embryos are homozygous for the specific mutation, it is documented as likely to be causal (Fig. 3d, e), with 10% tolerance for pipetting and genotyping errors. If less than 90% of embryos with one phenotype are homozygous for the mutation of interest, it is documented as being linked to a phenotype rather than causative.

We have carried out a phenotypic analysis of 1,216 nonsense and essential splice mutations. Of these, 48 mutations caused a phenotype within the first 5 d.p.f. and 77 alleles were linked to a phenotype. Among the predicted disruptive mutations, 1,065 were deemed to have no phenotype at 5 d.p.f. and 26 are under further investigation. For all phenotype–genotype correlations we annotated each of the phenotypic traits using developmental stage, anatomical entity and phenotypic quality terms to enable phenotype data mining.

This mutation discovery and the multi-allelic phenotyping pipeline systematically annotates zebrafish gene function. Importantly,

the described genotype and phenotype correlations do not constitute proof of causality for the individual allele. Detailed aetiology of a phenotype–genotype correlation can only be proven by more exhaustive investigation, such as a complementation test. By combining two distinct potentially disruptive mutations in the same gene using a compound cross of two independent heterozygous carriers and carrying out a genotype analysis of the expected single-phenotype F_3 embryos, we can rule out linked mutations independently associated with each allele. In cases in which a phenotype and allele have been associated and additional alleles in the same gene are available, we perform complementation crosses to prove causality (Fig. 4).

We find that approximately 6% (74 out of 1,216) of alleles are phenotypic, and this percentage is low in comparison to measurements of mouse embryonic lethality²³. There are several possible explanations. First, the alleles generated and analysed in this project are random point mutations across the length of each gene. Therefore, a proportion of alleles do not disrupt—or only partially disrupt—protein function. However, the position of a mutation is not necessarily a good predictor of the severity of disruption (demonstrated by the alleles described in Fig. 4d, h, k). Second, our phenotyping assays include only those morphological and behavioural changes that are detectable during the first 5 d.p.f. in live embryos. Subtle phenotypes that require further intervention, such as by immunohistochemistry, are not currently assayed. Finally, the teleost-specific genome duplication may

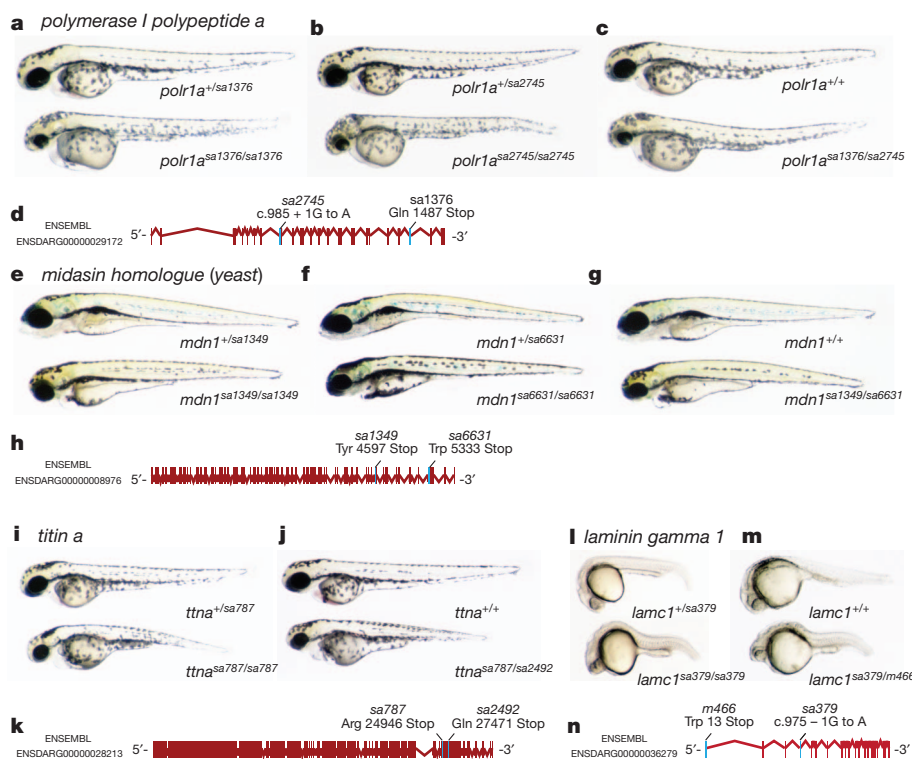


Figure 4 | Confirmation of causality through complementation crosses.

Four examples of complementation crosses, *polymerase (RNA) I polypeptide a* (*polr1a*) (a–d), *midasin homologue (yeast)* (*mdn1*) (e–h), *titin a* (*titna*) (i–l) and *laminin gamma 1* (*lamc1*) (m–n). Heterozygous carriers of two independent alleles in the same gene were used to generate compound heterozygote offspring. Incrosses of individual alleles, for which carriers of both sexes were available, are also shown. In all images, upper panels are non-phenotypic siblings and lower panels are phenotypic homozygous mutant or compound heterozygous embryos. a–c, At 48 hours post fertilization (h.p.f.) embryos homozygous for either *polr1a*^{sa1376} or *polr1a*^{sa2745}, or compound heterozygous for *polr1a*^{sa1376} and *polr1a*^{sa2745} have small eyes, a hydrocephalic hindbrain and pericardiac oedema. d, *polr1a*^{sa2745} disrupts a splice donor site through a G to A transition at the first intronic nucleotide 3' of coding nucleotide 985. Allele *polr1a*^{sa1376} is a C to T transition producing a premature stop codon at amino acid 1,487. e–g, At 96 h.p.f. embryos homozygous for either *mdn1*^{sa1349} or

mdn1^{sa6631}, or compound heterozygous for *mdn1*^{sa1349} and *mdn1*^{sa6631} have smaller heads with malformed jaws and mild pericardiac oedema. h, *mdn1*^{sa1349} and *mdn1*^{sa6631} produce premature stop codons at amino acid 4,597 (T to A transversion) and amino acid 5,333 (G to A transition), respectively. i, j, At 48 h.p.f. embryos homozygous for *titna*^{sa787} or compound heterozygous for *titna*^{sa787} and *titna*^{sa2492} are growth retarded, paralysed and have pericardiac oedema. k, Alleles *titna*^{sa787} and *titna*^{sa2492} produce premature stop codons at amino acid 24,946 (C to T transition) and amino acid 27,471 (C to T transition), respectively. l, m, At 24 h.p.f. embryos homozygous for *lamc1*^{sa379}, or compound heterozygous for *lamc1*^{sa379} and *lamc1*^{m466} are shorter with an undifferentiated notochord, and brain and eye malformations. n, Allele *lamc1*^{m466} is a G to A transition producing a premature stop codon at amino acid 13. Allele *lamc1*^{sa379} disrupts a splice acceptor site through a G to A transition one nucleotide 5' of coding nucleotide 975. Stop indicates a premature stop codon generated by a base change.

cause paralogue redundancy. Although this is possible, there are few examples of paralogues being completely redundant. In contrast, there are numerous paralogue pairs in which gene expression domains and functions have split between paralogues^{24,25}. For example, individuals homozygous for mutations in *titin a* (*tna*) (Fig. 4i–k) show a phenotype distinct from the published *tnb* mutant *runzel*²⁶.

It is unlikely that a comprehensive functional understanding of all human protein-coding genes will become available in the near future. Therefore, by providing a systematic analysis of zebrafish gene function, with phenotypes annotated in searchable, ontology-based data sets, the reagents described here will advance our knowledge of the biological basis for human disease. So far we have identified mutations in the orthologues of 3,188 of the 5,494 genes currently associated with human disease in genome-wide association studies (<http://www.genome.gov/gwastudies>) and have identified at least 1 allele in 2,505 of the 4,204 genes associated with a human phenotype in the Online Mendelian Inheritance in Man database (<http://www.omim.org>).

Our analysis will provide a rich resource for developmental biologists and clinicians to facilitate the identification of candidate genes for idiopathic inherited diseases or pathogen susceptibility. Furthermore, the alleles and data generated by this project are available to the scientific community through our website (http://www.sanger.ac.uk/Projects/D_rerio/zmp) and alleles will be available from two international stock centres, the Zebrafish International Resource Center (<http://www.zebrafish.org/zirc>) and the European Zebrafish Resource Center (<http://www.itg.kit.edu/ezrc>). Information on how to use these facilities can be found in the Supplementary Information. We believe that the work described here will substantially enhance the use of zebrafish as a model organism to study vertebrate development and human disease.

METHODS SUMMARY

Exon coordinates (Zv8 and later Zv9) were used to design the Agilent SureSelect baits. DNA was prepared from fin biopsies of F₁ progeny from mutagenized individuals, and barcoded Illumina sequencing libraries (150–200-bp insert size) were prepared and hybridized to the SureSelect baits. Exome-enriched libraries were amplified by PCR and subjected to 54-bp paired-end Illumina sequencing. Sequences were analysed using a custom computational pipeline to identify nonsense and essential splice mutations induced by the mutagenesis. Phenotyping was carried out by in-crossing up to 12 pairs of F₂ adults. Each breeding pair was fin clipped and genotyped for the induced nonsense and essential splice mutations that were detected in the F₁ exome sequence. From each breeding pair, 150 F₃ embryos were sorted into 3 dishes. Embryos were incubated at 28.5 °C and previous mutagenesis screens were used as a reference for the phenotyping^{27,28}. In the first round of the phenotyping, 48 phenotypically normal embryos were collected at 5 d.p.f. Embryos were then genotyped for the mutations that were heterozygous in both F₂ adults. In the second round, F₂ individuals that were heterozygous for a mutation suspected to cause a phenotype were re-crossed and the F₃ embryos were studied for phenotypes on all of the first 5 d.p.f. Phenotypic and non-phenotypic embryos were then genotyped for the mutation of interest. Genotyping was carried out using the KBioscience competitive allele-specific PCR (KASP) genotyping system.

Full Methods and any associated references are available in the online version of the paper.

Received 18 August 2012; accepted 7 February 2013.

Published online 17 April 2013.

- Gossler, A., Joyner, A. L., Rossant, J. & Skarnes, W. C. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* **244**, 463–465 (1989).
- Skarnes, W. C., Auerbach, B. A. & Joyner, A. L. A gene trap approach in mouse embryonic stem cells: the lacZ reported is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. *Genes Dev.* **6**, 903–918 (1992).
- Ringwald, M. *et al.* The IKMC web portal: a central point of entry to data and resources from the International Knockout Mouse Consortium. *Nucleic Acids Res.* **39**, D849–D855 (2011).
- Church, D. M. *et al.* Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biol.* **7**, e1000112 (2009).
- Chinwalla, A. T. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Zhang, J., Talbot, W. S. & Schier, A. F. Positional cloning identifies zebrafish one-eyed pinhead as a permissive EGF-related ligand required during gastrulation. *Cell* **92**, 241–251 (1998).
- Golling, G. *et al.* Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. *Nature Genet.* **31**, 135–140 (2002).
- Amsterdam, A. *et al.* Identification of 315 genes essential for early zebrafish development. *Proc. Natl Acad. Sci. USA* **101**, 12792–12797 (2004).
- Sun, Z. *et al.* A genetic screen in zebrafish identifies cilia genes as a principal cause of cystic kidney. *Development* **131**, 4085–4093 (2004).
- Nasevicius, A. & Ekker, S. C. Effective targeted gene 'knockdown' in zebrafish. *Nature Genet.* **26**, 216–220 (2000).
- Kettleborough, R. N., Bruijn, E., Eeden, F., Cuppen, E. & Stemple, D. L. High-throughput target-selected gene inactivation in zebrafish. *Methods Cell Biol.* **104**, 121–127 (2011).
- Sood, R. *et al.* Methods for reverse genetic screening in zebrafish by resequencing and TILLING. *Methods* **39**, 220–227 (2006).
- Wienholds, E. *et al.* Efficient target-selected mutagenesis in zebrafish. *Genome Res.* **13**, 2700–2707 (2003).
- Meng, X., Noyes, M. B., Zhu, L. J., Lawson, N. D. & Wolfe, S. A. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nature Biotechnol.* **26**, 695–701 (2008).
- Doyon, Y. *et al.* Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nature Biotechnol.* **26**, 702–708 (2008).
- Huang, P. *et al.* Heritable gene targeting in zebrafish using customized TALENs. *Nature Biotechnol.* **29**, 699–700 (2011).
- Sander, J. D. *et al.* Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nature Biotechnol.* **29**, 697–698 (2011).
- Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* (in the press).
- Collins, J. E., White, S., Searle, S. M. & Stemple, D. L. Incorporating RNA-seq data into the zebrafish ensembl gene build. *Genome Res.* **22**, 2067–2078 (2012).
- Stemple, D. L. TILLING—a high-throughput harvest for functional genomics. *Nature Rev. Genet.* **5**, 145–150 (2004).
- The 1000 Genome Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2011).
- The 1000 Genome Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
- Ayadi, A. *et al.* Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project. *Mamm. Genome* **23**, 600–610 (2012).
- Seiler, C. *et al.* Myosin VI is required for structural integrity of the apical surface of sensory hair cells in zebrafish. *Dev. Biol.* **272**, 328–338 (2004).
- Manfron, I. *et al.* Zebrafish *sox9b* is crucial for hepatopancreatic duct development and pancreatic endocrine cell regeneration. *Dev. Biol.* **366**, 268–278 (2012).
- Steffen, L. S. *et al.* The zebrafish *runzel* muscular dystrophy is linked to the *titin* gene. *Dev. Biol.* **309**, 180–192 (2007).
- Haffter, P. *et al.* The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* **123**, 1–36 (1996).
- Driever, W. *et al.* A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123**, 37–46 (1996).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Ellis, E. Markham, H. van Roekel, P. Toonen, J. van de Belt, J. Mudde and S. Widaa for technical assistance, and B. Novak, Y. Yi and E. LeProust from Agilent Technologies. We thank everyone at The Zebrafish International Resource Center and the European Zebrafish Resource Center for stocking and distributing alleles. We thank members of the Wellcome Trust Sanger Institute RSF and DNA pipelines. We also thank G. Powell, J. Collins and F. L. Marlow for critical reading of the manuscript. This work was funded through a core grant to the Sanger Institute by the Wellcome Trust (grant number 098051), the US National Institutes of Health (5R01HG004819), the EU Sixth Framework Programme (ZF-MODELS, contract number LSHG-CT-2003-503496) and the EU Seventh Framework Programme (ZF-HEALTH). F.v.E. is supported by the UK Medical Research Council (grant number G0777791) and E.C. is supported by the SmartMix program (SSM06010) from the Dutch government.

Author Contributions R.N.W.K., E.M.B.-N. and S.A.H. initiated, organized and executed the work (equal contributions). R.N.W.K., E.M.B.-N., S.A.H. and D.S. designed the experiments. R.N.W.K., E.M.B.-N. and S.A.H. wrote the manuscript with assistance from I.S., R.J.W., C.M.D. and D.L.S. Mutagenesis was carried out by R.N.W.K., F.v.E. and E.D.B. The mutation analysis pipeline was developed by I.S. and I.J.N., and maintained by R.J.W. C.H. and F.F. implemented and improved genotyping procedures. F.F. and E.M.B.-N. developed the cryopreservation procedure. S.A.H., S.M., C.S., C.M.D. and N.W. carried out the phenotyping. J.Y. designed and tested the first Agilent SureSelect exome set, R.G. helped to maintain and distribute alleles. S.C. and A.H. provided assistance for cryopreservation and genotyping. E.C. and D.L.S. collaborated in the initiation, design and process development of the project. All authors read the manuscript and provided comments.

Author Information Detailed information on alleles and their availability can be found online (http://www.sanger.ac.uk/Projects/D_rerio/zmp/). All sequencing data are deposited in the European Molecular Biology Laboratory (EMBL) European Nucleotide Archive under accession ERP000426. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.L.S. (ds4@sanger.ac.uk) or E.C. (e.cuppen@hubrecht.eu).

METHODS

Exome sequencing and SNP calling. Adult male zebrafish were mutagenized using ENU according to improved mutagenesis protocols¹¹. G₀ mutagenized individuals were outcrossed to create large F₁ mutagenized libraries. DNA was isolated from F₁ individuals by incubating fin biopsies in 400 µl of 100 µg ml⁻¹ proteinase K for 10 h at 55 °C, and it was then incubated for 15 min at 85 °C to heat-inactivate the proteinase K. DNA was precipitated by adding 400 µl of isopropanol and centrifuging for 30 min at 2,700g and 4 °C. DNA pellets were washed twice with 400 µl of 70% ethanol followed by centrifugation at 4,000 r.p.m. for 5 min, and re-suspended in ddH₂O. DNA from each individual (1–2 µg) was sheared and used to construct 150- to 200-bp-insert Illumina libraries according to the manufacturer's standard protocols.

For the rapid identification of ENU-induced mutations in individual zebrafish covering all annotated zebrafish protein-coding genes, we developed a whole-exome enrichment reagent using Agilent SureSelect. RNA oligonucleotides, 120 bases in length, were designed across the predicted exon coordinates to cover each base twice (2× tiling) and then manufactured as biotinylated RNA baits and blended into one tube, ready for enrichment. After completion of a pilot experiment to evaluate the technology, we developed an exome design using the Ensembl 61 (Zv9) gene set, which included a total of 60 Mbp of coding sequence and 26,206 genes^{18,19}.

For each F₁ genomic Illumina library, 500 ng of DNA was hybridized for 24 h to biotinylated whole exome RNA baits. Hybridized fragments were enriched using streptavidin-coated beads, RNA was digested, and remaining libraries of fragments were amplified for 10 cycles using standard Illumina primers with or without indexing barcodes. The resulting amplified libraries were run on Illumina GAI or HiSeq2000 machines using GAI, HiSeq2000v2 or HiSeq2000v3 chemistries to perform 54-bp paired-end sequencing.

Initially, each enriched sample was sequenced on an individual lane of an Illumina GAI machine using 54-bp paired-end runs, achieving a mean of 64 million reads per sample (Supplementary Table 1). Of those reads, 55% mapped to the exome target sequence, with 90% of the target being covered at 4× and 64% covered at 20×. For the identification of SNP variants, at least 4× coverage was required, with 20× coverage providing the number of reads required for reliable mutation detection²⁹. We subsequently moved to the HiSeq platform, incorporated barcoding into the production of the library and carried out pre-capture pooling of libraries. These improvements enabled us to sequence eight exomes on an individual lane, consequently increasing the throughput and lowering costs (Supplementary Table 1). These results show that we could efficiently enrich and sequence the zebrafish exome at the coverage required for reliable mutation detection in a cost-effective manner.

We identified ENU-induced mutations within the exome sequences using a modified version of the 1000 Genomes Project variant-calling pipeline²¹. Paired reads were aligned to the Zv9 reference assembly using the Burrows-Wheeler Aligner (BWA), and SNPs were called by SAMtools mpileup, QCALL and the GATK Unified Genotyper. SNPs that were not called by all three callers were removed from the analysis, along with any SNP that did not pass a caller's standard filters. In addition, SNPs were removed in cases in which the total read depth was less than the number of samples and if the genotype quality was lower than 100 for GATK and lower than 50 for QCALL and SAMtools mpileup. Finally, SNPs within 10 bp of an indel (called by both SAMtools mpileup and Dindel) were removed. Variation consequences were assigned by the Ensembl Variant Effect Predictor³⁰ using the Ensembl 61 gene set. An SNP was defined as an induced mutation if present in one to three individuals, as this allowed for founder effects that may have arisen from the mutagenesis. If the SNP was present in more than three individuals it was considered to be common to the strain used. Heterozygous calls that were found in one to three samples only were deemed to be induced mutations and those with a nonsense consequence (Sequence Ontology accession SO:0001587) or essential splice consequence (SO accessions: SO:0001574 and SO:0001575) were used to design KASP assays.

In a pilot run, using a sample of six individual exomes, we confirmed all induced mutations that were called by at least two of the SNP callers by both KASP genotyping and capillary sequencing to calculate the false positive rate of our analysis (data not shown). These genotyping results were used to set filtering parameters within our SNP calling pipeline, such that 85.7% of SNPs that could not be confirmed by KASP genotyping or capillary sequencing were removed. Information on common SNPs, and insertions and deletions, were also collected so that they could be avoided when designing genotyping assays.

To estimate the mutation saturation we used missense mutations, which are present in F₁ individuals at a tenfold higher rate than induced nonsense and essential splice-site mutations combined. As each mutagenized library displayed different rates of mutagenesis, the order that exomes were sequenced was randomized in the analysis.

We reasoned that the induced nonsense and essential splice mutations were more likely to result in putative loss-of-function alleles than non-synonymous mutations. Moreover, we did not include non-synonymous mutations for the phenotypic analysis on the assumption that if they are truly loss-of-function alleles, sequencing more individuals will eventually give a nonsense or essential splice allele in those genes.

Phenotyping. Zebrafish were maintained in accordance with UK Home Office regulations, UK Animals (Scientific Procedures) Act 1986, under project licence 80/2192, which was reviewed by the Wellcome Trust Sanger Institute Ethical Review Committee.

Heterozygous F₂ fish were randomly incrossed, and after egg collection F₂ adults had their fins clipped, were genotyped for the induced nonsense and essential splice mutations identified in the F₁ exome sequence, and kept as isolated breeding pairs. For each family we aimed to phenotype 12 pairs over 3 weeks of breeding. Each clutch of eggs, which was labelled with the identification numbers of the breeding pair, was sorted into three 10-cm Petri dishes containing approximately 50 embryos each. Embryos were incubated at 28.5 °C. Previous mutagenesis screens were used as a reference for the phenotyping^{27,28}. The phenotypes studied were: day 1, early patterning defects, early arrest, notochord, eye development, somites, patterning and cell death in the brain; day 2, cardiac defects, circulation of the blood, pigment (melanocytes), eye and brain development; day 3, cardiac defects, circulation of the blood, pigment (melanocytes), movement and hatching; day 4, cardiac defects, movement, pigment (melanocytes) and muscle defects; day 5, behaviour (hearing, balance, response to touch), swim bladder, pigment (melanocytes, xanthophores and iridophores), distribution of pigment, jaw, skull, axis length, body shape, notochord degeneration, digestive organs (intestinal folds, liver and pancreas), left-right patterning.

In the first round of phenotyping, all phenotypic embryos were discarded. At 5 d.p.f. more than 48 phenotypically wild-type embryos were collected. Embryos were fixed in 100% methanol and stored at -20 °C until genotyping for nonsense and essential splice mutations heterozygous in both parents was initiated. If 6 or more out of 48 embryos in the first round of genotyping were homozygous (corresponding to a *P* value of more than 0.05 in a chi-squared test and indicating no statistically significant difference compared to the expected 25% ratio) the allele was deemed not to cause a phenotype within the first 5 d.p.f. If the number of homozygous embryos was 5 or fewer (corresponding to a *P* value of less than 0.05 in a chi-squared test), then the allele was carried forward into the second round of phenotyping.

In the second round, F₂ individuals that were heterozygous for a suspected causal mutation were re-crossed to obtain clutches containing embryos homozygous for suspected causal mutations. All phenotypes observed in those clutches of embryos were counted, documented and photographed. Phenotypic embryos were fixed in 100% methanol and at 5 d.p.f. 48 wild-type phenotype embryos were also collected. All embryos were genotyped for the suspected causal mutations. An allele was documented as causing a phenotype if the phenotypic embryos were homozygous for the allele and the non-phenotypic sibling embryos were heterozygous or homozygous for the wild-type. For each allele we accepted up to 10% of phenotypic embryos to be not homozygous, to account for errors in egg collection. Such alleles were outcrossed for further genotyping with F₄ embryos later. If possible, alleles were also submitted to complementation tests.

Genotyping. Embryo DNA was prepared by first removing the embryos from 100% methanol, and individual embryos were then placed into wells of a 96-well plate. The well positions of phenotypic and non-phenotypic embryos were documented. Incubating the plate at 80 °C for 15 min evaporated all remaining methanol. DNA was extracted from embryos by incubating in 25 µl of lysis buffer (25 mM NaOH with 0.2 mM EDTA) at 95 °C for 30 min. After this, a volume of 25 µl of neutralization buffer (40 mM Tris-HCl) was added. Genotyping was carried out using the KASP genotyping system (KBioscience). Clipped fins and embryo DNA were diluted to a working concentration of 1.25–12.5 ng µl⁻¹. A volume of 4 µl of DNA was pipetted into black 384-well hard shell PCR plates and dried at 20 °C. When genotyping was carried out, the DNA was re-suspended by adding a 4-µl PCR mix, according to the manufacturer's protocol (KBioscience). PCR results were analysed using PHERAstar plus (BMG labtech) and the software KlusterCaller (KBioscience).

Cryopreservation of alleles. Sperm from individual males was collected by abdominal massage into 10-µl glass capillaries. The sperm were expelled into 245 µl 10% *N,N*-dimethylacetamide (DMA) in BSMIS (75 mM NaCl, 70 mM KCl, 2 mM CaCl₂, 1 mM MgSO₄, 20 mM Tris, pH 8.0) and mixed briefly by pipetting up and down. Eight aliquots of 35 µl each were pipetted directly into 2-ml cryovials and transferred immediately into a pre-chilled 50-ml Falcon tube on a dry ice and ethanol bath. After 30 min the samples were moved into liquid nitrogen for long-term storage. Storage location, date of collection and sample

quality were documented together with genotyping data for each archived male. Representative sperm samples were tested by *in vitro* fertilization.

For *in vitro* fertilization sperm samples were thawed by addition of 500 μ l 37 °C BSMIS to the frozen specimen. The sperm were then activated by the addition of 500 μ l 28 °C 0.5% fructose in egg water (0.018% (w/v) synthetic sea salt in reverse osmosis water) and mixed immediately with fresh eggs in a 6-cm glass Petri dish. Eggs were obtained by squeezing females and used within a few minutes of collection. Sperm motility and egg quality were monitored and documented. After

about 50 s the glass dish was filled with egg water and eggs were transferred into a 10-cm plastic Petri dish. Fertility rates were checked after incubating the eggs for a few hours at 28.5 °C.

29. Nielsen, R., Paul, J. S., Albrechtsen, A. & Song, Y. S. Genotype and SNP calling from next-generation sequencing data. *Nature Rev. Genet.* **12**, 443–451 (2011).
30. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).

The zebrafish reference genome sequence and its relationship to the human genome

Kerstin Howe^{1*}, Matthew D. Clark^{1,2*}, Carlos F. Torroja^{1,3}, James Torrance¹, Camille Berthelot^{4,5,6}, Matthieu Muffato⁷, John E. Collins¹, Sean Humphray^{1,8}, Karen McLaren¹, Lucy Matthews¹, Stuart McLaren¹, Ian Sealy¹, Mario Caccamo², Carol Churcher¹, Carol Scott¹, Jeffrey C. Barrett¹, Romke Koch⁹, Gerd-Jörg Rauch¹⁰, Simon White¹, William Chow¹, Britt Kilian¹, Leonor T. Quintais⁷, José A. Guerra-Assunção⁷, Yi Zhou¹¹, Yong Gu¹, Jennifer Yen¹, Jan-Hinnerk Vogel¹, Tina Eyre¹, Seth Redmond¹, Ruby Banerjee¹, Jianxiang Chi¹, Beiyuan Fu¹, Elizabeth Langley¹, Sean F. Maguire¹, Gavin K. Laird¹, David Lloyd¹, Emma Kenyon¹, Sarah Donaldson¹, Harminder Sehra¹, Jeff Almeida-King¹, Jane Loveland¹, Stephen Trevanion¹, Matt Jones¹, Mike Quail¹, Dave Willey¹, Adrienne Hunt¹, John Burton¹, Sarah Sims¹, Kirsten McLay¹, Bob Plumb¹, Joy Davis¹, Chris Clee¹, Karen Oliver¹, Richard Clark¹, Clare Riddle¹, David Elliott¹, Glen Threadgold¹, Glenn Harden¹, Darren Ware¹, Beverly Mortimer¹, Giselle Kerry¹, Paul Heath¹, Benjamin Phillimore¹, Alan Tracey¹, Nicole Corby¹, Matthew Dunn¹, Christopher Johnson¹, Jonathan Wood¹, Susan Clark¹, Sarah Pelan¹, Guy Griffiths¹, Michelle Smith¹, Rebecca Glithero¹, Philip Howden¹, Nicholas Barker¹, Christopher Stevens¹, Joanna Harley¹, Karen Holt¹, Georgios Panagiotidis¹, Jamieson Lovell¹, Helen Beasley¹, Carl Henderson¹, Daria Gordon¹, Katherine Auger¹, Deborah Wright¹, Joanna Collins¹, Claire Raisen¹, Lauren Dyer¹, Kenric Leung¹, Lauren Robertson¹, Kirsty Ambridge¹, Daniel Leongamornlert¹, Sarah McGuire¹, Ruth Gilderthorp¹, Coline Griffiths¹, Deepa Manthradadi¹, Sarah Nichol¹, Gary Barker¹, Siobhan Whitehead¹, Michael Kay¹, Jacqueline Brown¹, Clare Murnane¹, Emma Gray¹, Matthew Humphries¹, Neil Sycamore¹, Darren Barker¹, David Saunders¹, Justene Wallis¹, Anne Babbage¹, Sian Hammond¹, Maryam Mashreghi-Mohammadi¹, Lucy Barr¹, Sancha Martin¹, Paul Wray¹, Andrew Ellington¹, Nicholas Matthews¹, Matthew Ellwood¹, Rebecca Woodmansey¹, Graham Clark¹, James Cooper¹, Anthony Tromans¹, Darren Grafham¹, Carl Skuce¹, Richard Pandian¹, Robert Andrews¹, Elliot Harrison¹, Andrew Kimberley¹, Jane Garnett¹, Nigel Fosker¹, Rebekah Hall¹, Patrick Garner¹, Daniel Kelly¹, Christine Bird¹, Sophie Palmer¹, Ines Gehring¹⁰, Andrea Berger¹⁰, Christopher M. Dooley^{1,10}, Zübeyde Ersan-Ürün¹⁰, Cigdem Eser¹⁰, Horst Geiger¹⁰, Maria Geisler¹⁰, Lena Karotki¹⁰, Anette Kirn¹⁰, Judith Konantz¹⁰, Martina Konantz¹⁰, Martina Oberländer¹⁰, Silke Rudolph-Geiger¹⁰, Mathias Teucke¹⁰, Kazutoyo Osoegawa¹², Baoli Zhu¹², Amanda Rapp¹³, Sara Widaa¹, Cordelia Langford¹, Fengtang Yang¹, Nigel P. Carter¹, Jennifer Harrow¹, Zemin Ning¹, Javier Herrero⁷, Steve M. J. Searle¹, Anton Enright⁷, Robert Geisler^{10,14}, Ronald H. A. Plasterk⁹, Charles Lee¹⁵, Monte Westerfield¹³, Pieter J. de Jong¹², Leonard I. Zon¹¹, John H. Postlethwait¹³, Christiane Nüsslein-Volhard¹⁰, Tim J. P. Hubbard¹, Hugues Roest Crollius^{4,5,6}, Jane Rogers^{1,2} & Derek L. Stemple¹

Zebrafish have become a popular organism for the study of vertebrate gene function^{1,2}. The virtually transparent embryos of this species, and the ability to accelerate genetic studies by gene knock-down or overexpression, have led to the widespread use of zebrafish in the detailed investigation of vertebrate gene function and increasingly, the study of human genetic disease^{3–5}. However, for effective modelling of human genetic disease it is important to understand the extent to which zebrafish genes and gene structures are related to orthologous human genes. To examine this, we generated a high-quality sequence assembly of the zebrafish genome, made up of an overlapping set of completely sequenced large-insert clones that were ordered and oriented using a high-resolution high-density meiotic map. Detailed automatic and manual annotation provides evidence of more than 26,000 protein-coding genes⁶, the largest gene set of any vertebrate so far sequenced. Comparison to the human reference genome shows that approximately 70% of human genes have at least one obvious zebrafish orthologue. In addition, the high quality of this genome assembly provides a clearer understanding of key genomic features such as a unique repeat content, a scarcity of pseudogenes, an enrichment of zebrafish-specific genes on chromosome 4 and chromosomal regions that influence sex determination.

The zebrafish (*Danio rerio*) was first identified as a genetically tractable organism in the 1980s. The systematic application of genetic screens led to the phenotypic characterization of a large collection of mutations^{1,2}. These mutations, when driven to homozygosity, can produce defects in a variety of organ systems with pathologies similar to human disease. Such investigations have also contributed notably to our understanding of basic vertebrate biology and vertebrate development. In addition to enabling the systematic definition of a large range of early developmental phenotypes, screens in zebrafish have contributed more generally to our understanding of the factors controlling the specification of cell types, organ systems and body axes of vertebrates^{7–9}.

Although its contributions have already been substantial, zebrafish research holds further promise to enhance our understanding of the detailed roles of specific genes in human diseases, both rare and common. Increasingly, zebrafish experiments are included in studies of human genetic disease, often providing independent verification of the activity of a gene implicated in a human disease^{3,5,10}. Essential to this enterprise is a high-quality genome sequence and complete annotation of zebrafish protein-coding genes with identification of their human orthologues.

¹Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ²The Genome Analysis Centre, Norwich Research Park, Norwich NR4 7UH, UK. ³Bioinformatics Unit, Centro Nacional de Investigaciones Cardiovasculares, 28029 Madrid, Spain. ⁴Ecole Normale Supérieure, Institut de Biologie de l'ENS, IBENS, 46 rue d'Ulm, Paris F-75005, France. ⁵INSERM, U1024, 46 rue d'Ulm, Paris F-75005, France. ⁶CNRS, UMR 8197, 46 rue d'Ulm, Paris F-75005, France. ⁷EMBL European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. ⁸Illumina Cambridge, Chesterford Research Park, Little Chesterford, Saffron Walden CB10 1XL, UK. ⁹Hubrecht Laboratory, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands. ¹⁰Max Planck Institute for Developmental Biology, Spemannstraße 35, 72076 Tübingen, Germany. ¹¹Stem Cell Program and Division of Hematology and Oncology, Children's Hospital and Dana Farber Cancer Institute, 1 Blackfan Circle, Karp 7, Boston, Massachusetts 02115, USA. ¹²Children's Hospital Oakland, 747 52nd Street, Oakland, California 94609, USA. ¹³Institute of Neuroscience, University of Oregon, 1254 University of Oregon, 222 Huestis Hall, Eugene, Oregon 97403-1254, USA. ¹⁴Karlsruhe Institute of Technology (KIT), Campus North, Institute of Toxicology and Genetics (ITG), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany. ¹⁵Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.

Table 1 | Assembly and annotation statistics for the Zv9 assembly

Assembly		Annotation	
Total length (bp)	1,412,464,843	Protein-coding genes	26,206
Total clone length (bp)	1,175,673,296	Pseudogenes	218
Total WGS31 contig length (bp)	234,099,447	RNA genes	4,556
Placed scaffold length (bp)	1,357,051,643	Immunoglobulin/T-cell receptor gene segments	56
Unplaced scaffold length (bp)	55,413,200	Total transcripts	53,734
Maximum scaffold length (bp)	12,372,269	Total exons	323,599
Scaffold N50 (bp)	1,551,602	-	-
No. of clones	11,100	-	-
No of WGS31 contigs	26,199	-	-
No. of placed scaffolds	3,452	-	-
No. of unplaced scaffolds	1,107	-	-

Data are based on Ensembl version 67. N50, the scaffold size above which 50% of the total length of the sequence assembly can be found.

The zebrafish genome-sequencing project was initiated at the Wellcome Trust Sanger Institute in 2001. We chose Tübingen as the zebrafish reference strain as it had been used extensively to identify mutations affecting embryogenesis². Our strategy resembled the clone-by-clone sequencing approach adopted previously for both the human and mouse genome projects. The Zv9 assembly is a hybrid of high-quality finished clone sequence (83%) and whole-genome shotgun (WGS) sequence (17%), with a total size of 1.412 gigabases (Gb) (Table 1). The clone and WGS sequence is tied to a high-resolution, high-density meiotic map called the Sanger AB Tübingen map (SATmap), named after the strains of zebrafish used to make the map (Supplementary Information).

Zebrafish are members of the teleostei infraclass, a monophyletic group that is thought to have arisen approximately 340 million years ago from a common ancestor¹¹. Compared to other vertebrate species, this ancestor underwent an additional round of whole-genome duplication (WGD) called the teleost-specific genome duplication (TSD)¹². Gene duplicates that result from this process are called ohnologues (after Susumu Ohno who suggested this mechanism of gene duplication)¹³. Zebrafish possess 26,206 protein-coding genes⁶, more than any previously sequenced vertebrate, and they have a higher number of species-specific genes in their genome than do human, mouse or chicken. Some of this increased gene number is likely to be a consequence of the TSD.

A direct comparison of the zebrafish and human protein-coding genes reveals a number of interesting features. First, 71.4% of human genes have at least one zebrafish orthologue, as defined by Ensembl Compara¹⁴ (Table 2). Reciprocally, 69% of zebrafish genes have at least one human orthologue. Among the orthologous genes, 47% of human genes have a one-to-one relationship with a zebrafish orthologue. The second largest orthology class contains human genes that are associated with many zebrafish genes (the 'one-human-to-many-zebrafish' class), with an average of 2.28 zebrafish genes for each human gene, and this probably reflects the TSD. A few notable human genes have no clearly identifiable zebrafish orthologue; for example, the leukaemia inhibitory factor (*LIF*), oncostatin M (*OSM*) or interleukin-6 (*IL6*) genes, although the receptors *lifra*, *lifrb*, *osmr* and *il6r* are clearly present in the zebrafish genome. It is possible that zebrafish proteins with functionally similar activities to LIF, OSM and IL-6 exist, but that their sequence divergence is so great that they cannot be recognized as orthologues. Similarly, the zebrafish genome has no *BRCA1* orthologue, but does have an orthologue of the *BRCA1*-associated *BARD1* gene,

which encodes an associated and functionally similar protein and a *brca2* gene, which plays an important role in oocyte development, probably reflecting its role in DNA damage repair¹⁵.

Zebrafish have been used successfully to understand the biological activity of genes orthologous to human disease-related genes in greater detail^{3–5}. To investigate the number of potential disease-related genes, we compared the list of human genes possessing at least one zebrafish orthologue with the 3,176 genes bearing morbidity descriptions that are listed in the Online Mendelian Inheritance in Man (OMIM) database. Of these morbid genes, 2,601 (82%) can be related to at least one zebrafish orthologue. A similar comparison identified at least one zebrafish orthologue for 3,075 (76%) of the 4,023 human genes implicated in genome wide association studies (GWAS).

Zv9 shows an overall repeat content of 52.2%, the highest reported so far in a vertebrate. All other sequenced teleost fish exhibit a much lower repeat content, with an average of less than 30%. This result suggests that the evolutionary path leading to the zebrafish experienced an expansion of repeats, possibly facilitated by a population bottleneck. Alternatively, the repeat content of the other sequenced teleost species may be under-represented, as these assemblies are mostly WGS¹⁶.

The majority of transposable elements found in the human genome are type I (retrotransposable elements), with more than 4.3 million placements covering 44% of the sequence, whereas only 11% of the zebrafish genome sequence is covered by type I elements in less than 500,000 instances. In contrast, the zebrafish genome contains a marked excess of type II DNA transposable elements. Indeed, 2.3 million instances of type II DNA transposable elements cover 39% of the zebrafish genome sequence (Supplementary Table 12), whereas type II repeats cover only 3.2% of the human genome.

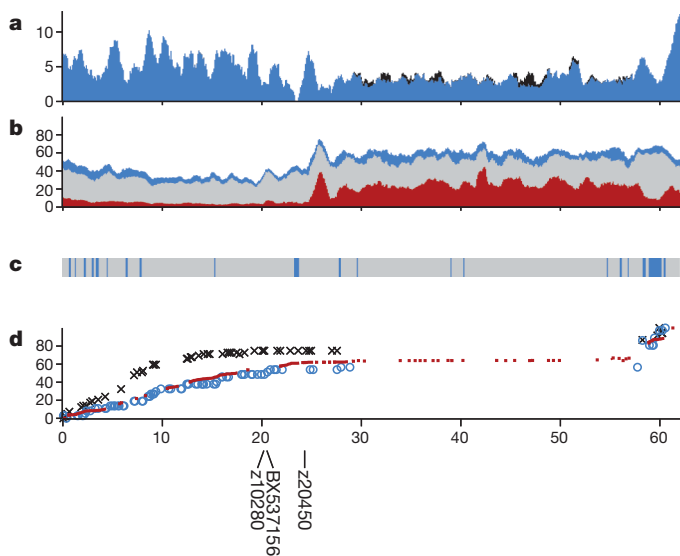
This pronounced abundance of type II transposable elements is unique among the sequenced vertebrate genomes, and the genome sequence shows evidence of recently active type II transposable elements. The closest vertebrate species in terms of the abundance of type II transposable elements is *Xenopus tropicalis* (25% type II transposable elements), whereas the sequenced and annotated teleost fish (the pufferfish *Takifugu* and *Tetraodon*, the three-spined stickleback (*Gasterosteus aculeatus*) and the medaka (*Oryzias latipes*)) each possess type II transposable element coverage of less than 10%, which may relate to the fact that the zebrafish genome diverges basally from the other sequenced and annotated teleost genomes¹⁷. Zebrafish type II transposable elements are divided into 14 superfamilies with 401 repeat families in total (Supplementary Table 12). The DNA and hAT superfamilies are the most abundant and diverse in the zebrafish genome, together covering 28% of the sequence. The type II transposable element abundance of zebrafish, or lack of retrotransposable elements, may provide an explanation for the low zebrafish pseudogene content (Supplementary Table 14).

The long arm of chromosome 4 is unique among zebrafish genomic regions, owing to its relative lack of protein-coding genes and its extensive heterochromatin. Chromosome 4 is known to be late-replicating and hybridization studies suggest that genomic copies of 5S ribosomal DNA (rDNA), which are not notably present on any other chromosome, are scattered along the long arm at high redundancy¹⁸.

Table 2 | Comparison of human and zebrafish protein-coding genes and their orthology relationships

Relationship type	Human	Core relationship	Zebrafish	Ratio
One to one	-	9,528	-	-
One to many	3,105	-	7,078	1:2.28
Many to one	1,247	-	489	2.55:1
Many to many	743	233	934	1:1.26
Orthologous total	14,623	13,355	18,029	1:1.28
Unique	5,856	-	8,177	-
Coding-gene total	20,479	-	26,206	-

Data and orthology relationship definitions are based on Ensembl Compara version 67 (http://www.ensembl.org/info/docs/compara/homology_method.html).



Immediately after the presumed centromere at approximately 24 megabases (Mb), the sequence landscape (Fig. 1 and Supplementary Fig. A4) shows a remarkable increase in repeat content, which continues through to the telomere of the long arm. At approximately 27 Mb,

Figure 1 | Landscape of chromosome 4. **a**, Exon coverage (blue), stacked with coverage by snRNA exons (black). **b**, Stacked repeat coverage, divided into type I transposable elements (red), type II transposable elements (grey) and other repeat types (blue), including dust, tandem and satellite repeats. **c**, Sequence composition (grey bars, clones; blue bars, WGS contigs). **d**, Genetic marker placements (red, SATmap markers; blue, heat shock meiotic map markers; black, Massachusetts General Hospital meiotic map markers). Marker placements have been normalized so that the maps can be compared. Near-centromeric clones are positioned at 20 Mb (BX537156), 20.2 Mb (Z10280) and 24.4 Mb (Z20450)²⁸. The *x* axis shows the chromosomal position in Mb. **a** and **b** were calculated as percentage coverage over 1-Mb overlapping windows (y axis), with a 100-kb shift between each window. **c** and **d** were calculated over 100-kb windows. The *y* axis for **d** shows the normalization of marker positions relative to the span of the individual map. Similar graphs for the other chromosome are provided in the Supplementary Information.

the otherwise uniform presence of the satellite repeat SAT-2 on the long arm ends abruptly. This location is also the starting point of uniform MOSAT-2 distribution, a satellite repeat that is nearly absent from all other chromosomes but highly enriched on the long arm of chromosome 4. The subtelomeric region of the long arm shows a distinct distribution of repeat elements, with relatively fewer interspersed elements and an increased content of satellite, simple and tandem repeats that do not harbour 5S rDNA sequences. Moreover, the gene content is reduced on the long arm and the guanine–cytosine content is slightly increased.

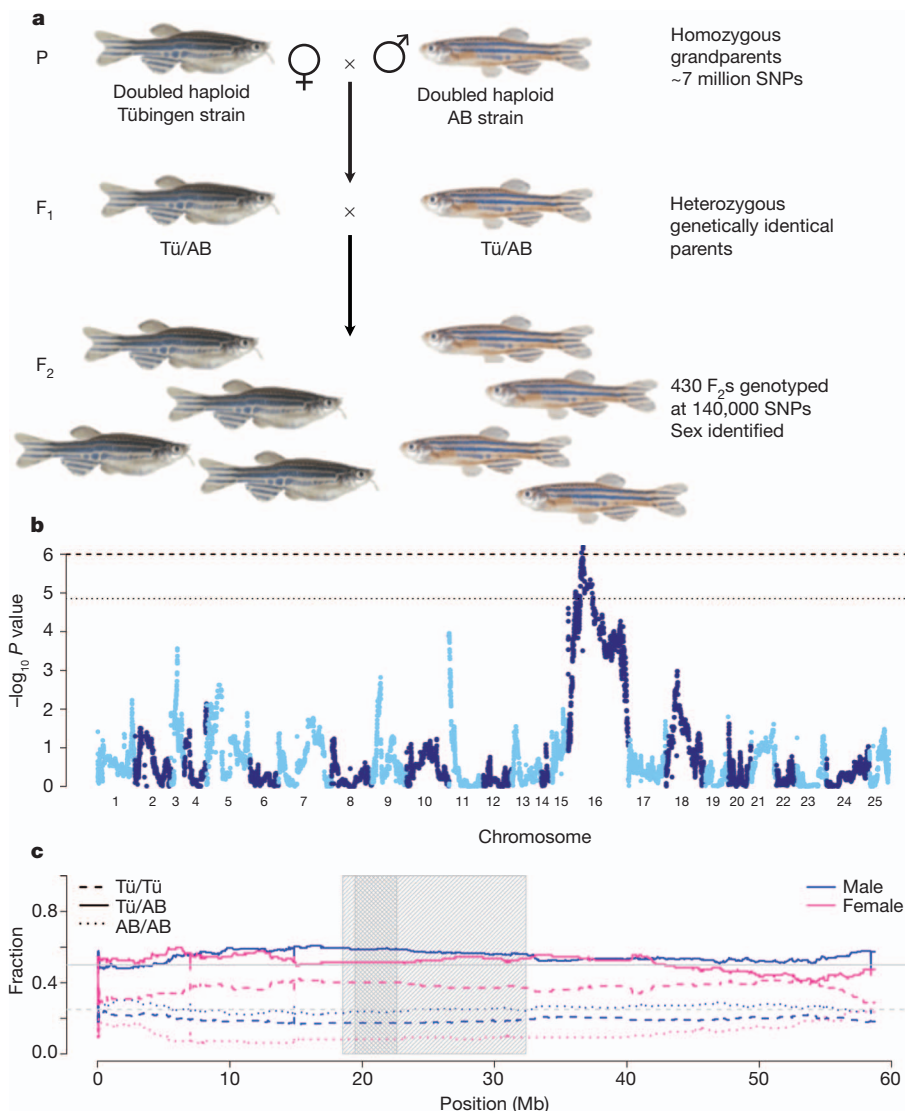


Figure 2 | Sex determination signal on chromosome 16. **a**, Breeding scheme for SATmap. Double haploid generation zero (G_0) founders were sequenced to approximately 40× depth using Illumina GAI technology. We found approximately 7 million SNPs between the two SATmap founders. This number of SNPs between just two homozygous zebrafish individuals is far in excess of that seen between any two humans and is nearly one-fifth of all SNPs measured among 1,092 human diploid genomes²⁹. Genetically identical, heterozygous F_1 fish of both sexes resulted from crossing the founders. The F_1 individuals were crossed to generate a panel of F_2 individuals, each with its own unique set of meiotic recombinations between AB and Tübingen (Tü) chromosomes, which were uncovered by dense genotyping with a set of 140,306 SNPs covering most of the genome. **b**, Genome-wide P values for tests of genotype difference between sexes, arranged by chromosome. The dotted line corresponds to differences that are expected once in 100 random genome scans, and the dashed line corresponds to differences expected once in 1,000 random genome scans. The only locus that is statistically significant at these levels is on chromosome 16. **c**, Genotype frequencies for males and females on chromosome 16. The grey line at 0.5 corresponds to expectation for heterozygotes (solid lines) and the grey line at 0.25 corresponds to expectation for homozygotes (dashed and dotted lines). The light grey shaded box corresponds to the region in which empirical $P < 0.01$, the dark grey shaded box corresponds to the region in which $P < 0.001$.

The long arm of chromosome 4 also has a special structure with respect to gene orthology and synteny. Approximately 80% of the genes present have no identifiable orthologues in human. In fact, 110 genes (out of 663) have no identifiable orthologues in any other sequenced teleost genome and indeed seem to be zebrafish-specific genes. The genes in this region are highly duplicated, with 31 ancestral gene families alone providing 77.5% of the genes, the largest of which contains no less than 109 duplicates in this region. The largest of these families correspond to NOD-like receptor proteins¹⁹ with putative roles in innate immunity and zinc finger proteins. We also observed a very high density of small nuclear RNAs (snRNAs) on chromosome 4, and in particular those that encode spliceosome components. The cohort of snRNAs carried on the long arm of chromosome 4 accounts for 53.2% of all snRNAs in the zebrafish genome. In addition, in a specific group of zebrafish derived recently from a natural population, the subtelomeric region of the long arm of chromosome 4 has been found to contain a major sex determinant with alleles that are 100% predictive of male development and 85% predictive of female development, suggesting that this chromosome may be, might have been, or may be becoming, a sex chromosome in this particular population²⁰.

In addition to the chromosome 4 sex determinant, three other separate genomic regions have been identified as influencing sex determination, and these vary between the strains and even within the families studied^{20,21}. Our meiotic map, SATmap, which was generated to anchor the genomic sequence, provided an opportunity to examine whether there are any strong signals for sex determination. To generate SATmap we took advantage of the fact that it is possible to create double haploid individuals that contain only maternally derived DNA, that are homozygous at every locus and that can be raised until they are fertile²² (Fig. 2a). To investigate the interesting finding that SATmap F₁ fish could be either male or female while being genetically identical and heterozygous at every polymorphic locus, we sought a genetic signal for sex determination in the F₂ generation, in which these polymorphisms segregate. Using morphological secondary sexual traits, we were able to score the sex of 332 genotyped F₂ individuals. Although most chromosomes showed no significant genetic bias for a particular sex, we found that most of chromosome 16 carried a strong signal ($P = 9.1 \times 10^{-7}$) with a broad peak around the centromere (Fig. 2b, c). Homozygotes for the Tübingen (grandmaternal) allele had a very high probability of being female, whereas homozygotes for the AB (grandpaternal) allele were very unlikely to be female (Fig. 2).

The number of protein-coding genes among vertebrates is relatively stable, although even closely related species may show great disparities in the nature of their protein-coding gene content. We carried out a four-way comparison between the proteome of two mammals (human and mouse), a bird (chicken) and the zebrafish to quantify the fraction of shared and species-specific genes present in each genome (Fig. 3a). A core group of 10,660 genes is found in all four species and probably approximates an essential set of vertebrate protein-coding genes. This number is somewhat less than the core set of 11,809 vertebrate genes identified previously as being common to three fish genomes (*Tetraodon*, medaka, zebrafish) and three amniotes (human, mouse, chicken)¹⁶, but the discrepancy probably reflects the improved annotation of these genomes that often results in fusing fragmented gene structures. Each taxon has between 2,596 and 3,634 species-specific genes. The notable excess observed in zebrafish may be a consequence of the WGD, because pairs of duplicated genes that arose from the WGD, but with no orthologue in amniotes, are counted as two specific genes. Furthermore, 2,059 genes are found in human, mouse and zebrafish but not in chicken, and this number is two times higher than the number of genes that are found in all amniotes but not in zebrafish (892). It is unclear whether these genes have been lost along the evolutionary branch leading to the chicken, or whether this is due to annotation or orthology assignment errors in the chicken genome.

We identified double-conserved synteny (DCS) blocks between all sequenced tetrapods and four fish genomes (zebrafish, medaka, stickleback and *Tetraodon*). DCS blocks are defined as runs of genes in the non-duplicated species that are found on two different chromosomes in the species that underwent a WGD²³, although the genes may not be adjacent in the duplicated species²⁴. The DCS between zebrafish and human are represented on either side of each human chromosome (Supplementary Fig. 15). Using DCS blocks, we identified zebrafish paralogous genes that are part of DCS blocks and consistent with the locally alternating chromosomes, hence with an origin at the TSD. We identified 3,440 pairs of such ohnologues (26% of the all genes), for a total of 8,083 genes when subsequent duplications are taken into account. It is notable that although true pairs of ohnologues may exist within the same chromosome owing to post-TSD rearrangements, we excluded such cases as we cannot reliably distinguish them from segmental duplications. This number of ancestral genes retained as

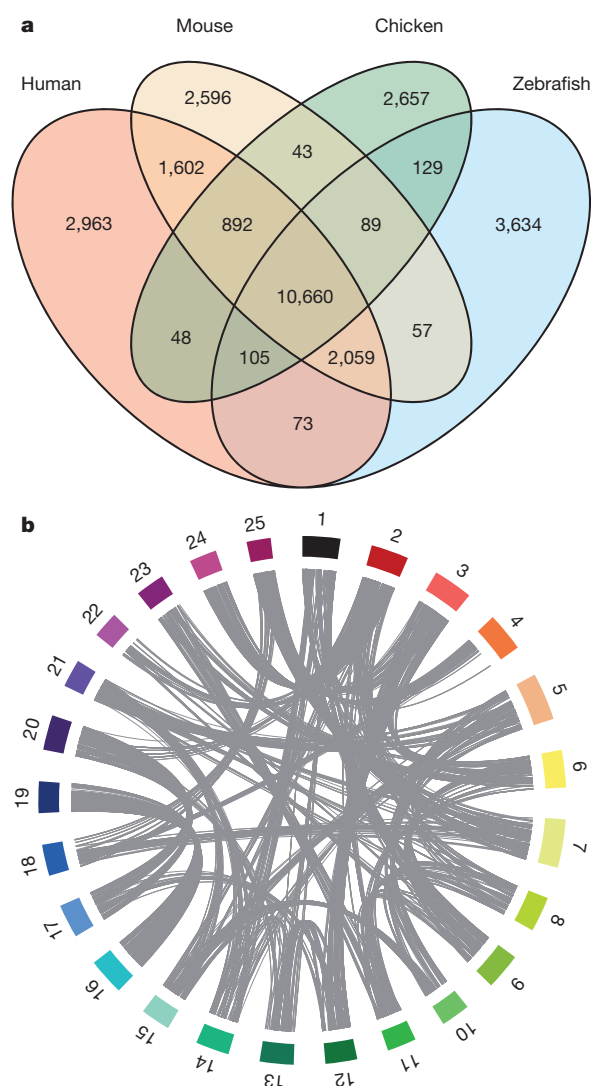


Figure 3 | Evolutionary aspects of the zebrafish genome. **a**, Orthologue genes shared between the zebrafish, human, mouse and chicken genomes, using orthology relationships from Ensembl Compara 63. Genes shared across species are considered in terms of copies at the time of the split. For example, a gene that exists in one copy in zebrafish but has been duplicated in the human lineage will be counted as only one shared gene in the overlap. **b**, The orthology relationships between zebrafish chromosomes. Chromosomes are represented as coloured blocks. The position of ohnologous genes between chromosomes are linked in grey (for clarity, links between chromosomes that share less than 20 ohnologues have been omitted). The image was produced using Circos³⁰.

duplicates in zebrafish is higher, both in absolute number and in proportion, than in other fish genomes (chi-squared test, all $P < 3 \times 10^{-5}$).

We compared the 8,083 zebrafish TSD ohnologues with human ohnologues originating from the two rounds of WGD that are common to all vertebrates and find that the two sets overlap strongly (chi-squared test, $P < 2 \times 10^{-16}$). In general, zebrafish ohnologous pairs are enriched in specific functions (neural activity, transcription factors) and are orthologous to mammalian genes under stronger evolutionary constraint than genes that have lost their second copy.

A circular representation of ohnologue pairs (Fig. 3b) highlights chromosomes, or parts of chromosomes, that descended from the same pre-duplication ancestral chromosome (for example, chromosomes 3 and 12, 17 and 20, 16 and 19). Among zebrafish chromosomes, chromosome 16 and chromosome 19 are unique in their one-to-one conservation of synteny. Consistent with the conservation of synteny, chromosome 16 and chromosome 19 possess clusters of orthologues of genes associated with the mammalian major histocompatibility complex (MHC) as well as the *hoxab* and *hoxaa* clusters, respectively, which are each orthologous to the human *HOXA* cluster²⁵.

Since the earliest whole-genome shotgun-only assembly became public in 2002, the zebrafish reference genome sequence has enabled many new discoveries to be made, in particular the positional cloning of hundreds of genes from mutations affecting embryogenesis, behaviour, physiology, and health and disease. Moreover, the annotated reference genome has enabled the generation of accurate whole-exome enrichment reagents, which are accelerating both positional cloning projects and new genome-wide mutation discovery efforts^{26,27}. Although the zebrafish reference genome sequencing is complete, a few poorly assembled regions remain, which are being resolved by the Genome Reference Consortium (<http://genomereference.org>).

METHODS SUMMARY

We generated cloned libraries of large fragments of genomic DNA, assembled a physical map of large-insert clones and completely sequenced a set of minimally overlapping clones. In addition, we generated WGS sequences by end-sequencing a mixture of large- and short-insert libraries. Overlapping clone sequences were combined with WGS sequences and tied to the meiotic map, SATmap, which enabled independent placement and orientation of clones in the genome sequence. The sequence data can be found in the BioProject database, under accession number PRJNA11776.

To obtain evidence for a more complete description of protein-coding genes, we used high-throughput short-read complementary DNA sequencing and obtained a deep-coverage data set for messenger RNAs expressed in zebrafish at various stages of development and in adult tissues⁶. Finally, a standard Ensembl gene build, incorporating filtered elements from the complementary DNA sequencing gene build, was merged with the manually curated gene models to produce a comprehensive annotation in Ensembl version 67 (http://may2012.archive.ensembl.org/Danio_rerio/Info/Index). Detailed descriptions of all the methods used for this project are available in the Supplementary Information.

Received 23 August 2012; accepted 21 March 2013.

Published online 17 April 2013.

1. Driever, W. *et al.* A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123**, 37–46 (1996).
2. Haffter, P. *et al.* The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* **123**, 1–36 (1996).
3. Golzio, C. *et al.* KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* **485**, 363–367 (2012).
4. Panizzi, J. R. *et al.* CCDC103 mutations cause primary ciliary dyskinesia by disrupting assembly of ciliary dynein arms. *Nature Genet.* **44**, 714–719 (2012).
5. Roscioli, T. *et al.* Mutations in *ISPD* cause Walker-Warburg syndrome and defective glycosylation of alpha-dystroglycan. *Nature Genet.* **44**, 581–585 (2012).
6. Collins, J. E., White, S., Searle, S. M. & Stemple, D. L. Incorporating RNA-seq data into the zebrafish Ensembl genebuild. *Genome Res.* **22**, 2067–2078 (2012).
7. Talbot, W. S. *et al.* A homeobox gene essential for zebrafish notochord development. *Nature* **378**, 150–157 (1995).
8. Gritsman, K. *et al.* The EGF-CFC protein one-eyed pinhead is essential for nodal signaling. *Cell* **97**, 121–132 (1999).
9. Ober, E. A., Verkade, H., Field, H. A. & Stainier, D. Y. Mesodermal Wnt2b signalling positively regulates liver specification. *Nature* **442**, 688–691 (2006).

10. Tobin, D. M. *et al.* Host genotype-specific therapies can optimize the inflammatory response to mycobacterial infections. *Cell* **148**, 434–446 (2012).
11. Amores, A., Catchen, J., Ferrara, A., Fontenot, Q. & Postlethwait, J. H. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics* **188**, 799–808 (2011).
12. Meyer, A. & Schartl, M. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin. Cell Biol.* **11**, 699–704 (1999).
13. Wolfe, K. Robustness—it's not where you think it is. *Nature Genet.* **25**, 3–4 (2000).
14. Vilella, A. J. *et al.* EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* **19**, 327–335 (2009).
15. Rodríguez-Mari, A. *et al.* Roles of *brca2* (*fancd1*) in oocyte nuclear architecture, gametogenesis, gonad tumors, and genome stability in zebrafish. *PLoS Genet.* **7**, e1001357 (2011).
16. Kasahara, M. *et al.* The medaka draft genome and insights into vertebrate genome evolution. *Nature* **447**, 714–719 (2007).
17. Postlethwait, J. H. The zebrafish genome in context: ohnologs gone missing. *J. Exp. Zool. B* **308**, 563–577 (2007).
18. Sola, L. & Gornung, E. Classical and molecular cytogenetics of the zebrafish, *Danio rerio* (Cyprinidae, Cypriniformes): an overview. *Genetica* **111**, 397–412 (2001).
19. Stein, C., Caccamo, M., Laird, G. & Leptin, M. Conservation and divergence of gene families encoding components of innate immune response systems in zebrafish. *Genome Biol.* **8**, R251 (2007).
20. Anderson, J. L. *et al.* Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. *PLoS ONE* **7**, e40701 (2012).
21. Bradley, K. M. *et al.* An SNP-based linkage map for zebrafish reveals sex determination loci. *G3 (Bethesda)* **1**, 3–9 (2011).
22. Streisinger, G., Walker, C., Dower, N., Knauber, D. & Singer, F. Production of clones of homozygous diploid zebra fish (*Brachydanio rerio*). *Nature* **291**, 293–296 (1981).
23. Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617–624 (2004).
24. Jaillon, O. *et al.* Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **431**, 946–957 (2004).
25. Amores, A. *et al.* Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish. *Genome Res.* **14**, 1–10 (2004).
26. Kettleborough, R. N. W. *et al.* A systematic genome-wide analysis of zebrafish protein-coding gene function. *Nature* (in the press).
27. Varshney, G. K. *et al.* A large-scale zebrafish gene knockout resource for the genome-wide study of gene function. *Genome Res.* **23**, 727–735 (2013).
28. Freeman, J. L. *et al.* Definition of the zebrafish genome using flow cytometry and cytogenetic mapping. *BMC Genomics* **8**, 195 (2007).
29. The 1000 Genomes Project Consortium An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
30. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We wish to thank R. Durbin, E. Birney, A. Scally, C. P. Ponting, E. Busch-Nentwich and R. Kettleborough for helpful discussions, as well as F. L. Marlow and P. Aanstad for critical reading and helpful comments on manuscripts. We thank the zebrafish information network (ZFIN) for funding part of the manual annotation of the zebrafish genome and the ZFIN staff for support with gene nomenclature and other genome issues. We also thank the Genome Reference Consortium for the maintenance and improvement of the zebrafish genome assembly. We are indebted to the Ensembl team for providing a browser and database that greatly facilitated the use and the analyses of the zebrafish genome. We thank A. Pirani at Affymetrix for genotyping advice support, and the Zebrafish International Resource Center (ZIRC) for distributing the SAT strain. J.H.P. was supported by the National Institutes of Health (NIH) grant R01 GM085318 (to J.H.P.), NIH grant P01 HD22486 (to J.H.P.) and R01 OD011116 (later changed to R01 RR020833) (to J.H.P.). We would like to acknowledge the support of the European Commission's Sixth Framework Programme (contract no. LSHG-CT-2003-503496, ZF-MODELS) and Seventh Framework Programme (grant no. HEALTH-F4-2010-242048, ZF-HEALTH). R.G. was supported by the German Human Genome Project (DHGP Grant 01 KW 9627 and 01 KW 9919). C.N.-V., G.-J.R. and R.G. were supported by the NIH (NIH grant 1 R01 DK55377-01A1). The Zebrafish Genome Project at the Wellcome Trust Sanger Institute was funded by Wellcome Trust grant number 098051.

Author Contributions K.H., M.D.C., D.L.S., C.B., H.R.C., A.E. and K.M. wrote the manuscript and Supplementary Information. M.D.C., C.F.T., I.S., J.C.B., A.R., S.W. and C.L. produced the SATmap. Z.N. and Y.G. produced the WGS31 assembly. J.T., W.C. and C.F.T. generated the Zv9 assembly. Previous assemblies were produced by M.C., who developed the first assembly integration process, and by S.R., T.E. and I.S. coordinated by K.H. The analyses and figures for the manuscript were produced by J.T., K.H., C.B., M.M., J.H., L.T.Q., J.A.G.-A. and J.Y. K.A., J.W., S.P., J.C., G.T., G.H., G.G., P.H. and B.K. are involved in the ongoing improvement of the zebrafish genome assembly. Manual annotation was produced by G.K.L., D.L., E.K., S.D., H.S., J.A.-K. and J.L. and coordinated by J.H. and M.W. Automated annotation (Ensembl) was provided by J.E.C., S.W., J.-H.V., S.T. and S.M.J.S. The genome sequencing was carried out by C.C., K.M., S.M., C.S., J.C., B.F., E.L., S.F.M., M.J., M.Q., D.W., A.H., J.B., S.S., K.M., B.P., J.D., C.C., K.O., B.M., G.K., B.P., A.T., N.C., C.J., S.C., M.S., R.G., P.H., N.B., C.S., J.H., K.H., G.P., J.L., H.B., C.H., D.G., D.W., C.R., L.D., K.L., L.R., K.A., D.L., S.M., R.G., C.G., D.M., S.N., G.B., S.W., M.K., J.B., C.M., E.G., M.H., N.S., D.B., D.S., J.W., A.B., S.H., K.O., M.M.-M., L.B., S.M., P.W., A.E., N.M., M.E., R.W., G.C., J.C.,

A.T., D.G., C.S., R.P., R.A., E.H., A.K., J.G., N.F., R.H., P.G., D.K., C.B. and S.P. The generation of maps used in the initial assemblies and the production of clone tiling paths were carried out by R.K., S.H., G.-J.R., Y.Z., C.R., R.C., D.E., D.W., L.M., M.D., I.G., A.B., C.M.D., Z.E.-Ü., C.E., H.G., M.G., L.K., A.K., J.K., M.K., M.O., S.R.-G., M.T., R.B., F.Y., N.P.C., R.G., R.H.A.P. and C.L. K.O., B.Z. and P.J.d.J. generated and provided clone libraries. The Zebrafish Genome Project was coordinated by L.I.Z., J.H.P., C.N.-V., T.J.P.H., J.R. and D.L.S.

Author Information Sequence data have been submitted to the BioProject database under accession PRJNA11776. Reprints and permissions information is available at

www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.L.S. (ds4@sanger.ac.uk).



This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

The global distribution and burden of dengue

Samir Bhatt¹, Peter W. Gething¹, Oliver J. Brady^{1,2}, Jane P. Messina¹, Andrew W. Farlow¹, Catherine L. Moyes¹, John M. Drake^{1,3}, John S. Brownstein⁴, Anne G. Hoen⁵, Osman Sankoh^{6,7,8}, Monica F. Myers¹, Dylan B. George⁹, Thomas Jaenisch¹⁰, G. R. William Wint^{1,11}, Cameron P. Simmons^{12,13}, Thomas W. Scott^{9,14}, Jeremy J. Farrar^{12,13,15} & Simon I. Hay^{1,9}

Dengue is a systemic viral infection transmitted between humans by *Aedes* mosquitoes¹. For some patients, dengue is a life-threatening illness². There are currently no licensed vaccines or specific therapeutics, and substantial vector control efforts have not stopped its rapid emergence and global spread³. The contemporary worldwide distribution of the risk of dengue virus infection⁴ and its public health burden are poorly known^{2,5}. Here we undertake an exhaustive assembly of known records of dengue occurrence worldwide, and use a formal modelling framework to map the global distribution of dengue risk. We then pair the resulting risk map with detailed longitudinal information from dengue cohort studies and population surfaces to infer the public health burden of dengue in 2010. We predict dengue to be ubiquitous throughout the tropics, with local spatial variations in risk influenced strongly by rainfall, temperature and the degree of urbanization. Using cartographic approaches, we estimate there to be 390 million (95% credible interval 284–528) dengue infections per year, of which 96 million (67–136) manifest apparently (any level of disease severity). This infection total is more than three times the dengue burden estimate of the World Health Organization². Stratification of our estimates by country allows comparison with national dengue reporting, after taking into account the probability of an apparent infection being formally reported. The most notable differences are discussed. These new risk maps and infection estimates provide novel insights into the global, regional and national public health burden imposed by dengue. We anticipate that they will provide a starting point for a wider discussion about the global impact of this disease and will help to guide improvements in disease control strategies using vaccine, drug and vector control methods, and in their economic evaluation.

Dengue is an acute systemic viral disease that has established itself globally in both endemic and epidemic transmission cycles. Dengue virus infection in humans is often inapparent^{1,6} but can lead to a wide range of clinical manifestations, from mild fever to potentially fatal dengue shock syndrome². The lifelong immunity developed after infection with one of the four virus types is type-specific¹, and progression to more serious disease is frequently, but not exclusively, associated with secondary infection by heterologous types^{2,5}. No effective antiviral agents yet exist to treat dengue infection and treatment therefore remains supportive². Furthermore, no licensed vaccine against dengue infection is available, and the most advanced dengue vaccine candidate did not meet expectations in a recent large trial^{7,8}. Current efforts to curb dengue transmission focus on the vector, using combinations of chemical and biological targeting of *Aedes* mosquitoes and management of breeding sites². These control efforts have failed to stem the increasing incidence of dengue fever epidemics and expansion of the

geographical range of endemic transmission⁹. Although the historical expansion of this disease is well documented, the potentially large burden of ill-health attributable to dengue across much of the tropical and subtropical world remains poorly enumerated.

Knowledge of the geographical distribution and burden of dengue is essential for understanding its contribution to global morbidity and mortality burdens, in determining how to allocate optimally the limited resources available for dengue control, and in evaluating the impact of such activities internationally. Additionally, estimates of both apparent and inapparent infection distributions form a key requirement for assessing clinical surveillance and for scoping reliably future vaccine demand and delivery strategies. Previous maps of dengue risk have used various approaches combining historical occurrence records and expert opinion to demarcate areas at endemic risk^{10–12}. More sophisticated risk-mapping techniques have also been implemented^{13,14}, but the empirical evidence base has since been improved, alongside advances in disease modelling approaches. Furthermore, no studies have used a continuous global risk map as the foundation for dengue burden estimation.

The first global estimates of total dengue virus infections were based on an assumed constant annual infection rate among a crude approximation of the population at risk (10% in 1 billion (ref. 5) or 4% in 2 billion (ref. 15)), yielding figures of 80–100 million infections per year worldwide in 1988 (refs 5, 15). As more information was collated on the ratio of dengue haemorrhagic fever to dengue fever cases, and the ratio of deaths to dengue haemorrhagic fever cases, the global figure was revised to 50–100 million infections^{16,17}, although larger estimates of 100–200 million have also been made¹⁰ (Fig. 1). These estimates were intended solely as approximations but, in the absence of better evidence, the resulting figure of 50–100 million infections per year is widely cited and currently used by the World Health Organization (WHO). As the methods used were informal, these estimates were presented without confidence intervals, and no attempt was made to assess geographical or temporal variation in incidence or the inapparent infection reservoir.

Here we present the outcome of a new project to derive an evidence-based map of dengue risk and estimates of apparent and inapparent infections worldwide on the basis of the global population in 2010. We compiled a database of 8,309 geo-located records of dengue occurrence from a systematic search, resulting from 2,838 published literature sources as well as newer online resources¹⁸ (see Supplementary Information, section A; the full bibliography⁴ and occurrence data are available from authors on request). Using these occurrence records we chose a set of gridded environmental and socioeconomic covariates known, or proposed, to affect dengue transmission (see Supplementary Information, section B); incorporated recent work assessing the strength of evidence on national and subnational-level dengue

¹Spatial Ecology and Epidemiology Group, Tinbergen Building, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. ²Oxitec Limited, Milton Park, Abingdon OX14 4RX, UK. ³Odom School of Ecology, University of Georgia, Athens, Georgia 30602, USA. ⁴Department of Pediatrics, Harvard Medical School and Children's Hospital Informatics Program, Boston Children's Hospital, Boston, Massachusetts 02115, USA. ⁵Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire 03755, USA. ⁶INDEPTH Network Secretariat, East Legon, PO Box KD 213, Accra, Ghana. ⁷School of Public Health, University of the Witwatersrand, Braamfontein 2000, Johannesburg, South Africa. ⁸Institute of Public Health, University of Heidelberg, 69120 Heidelberg, Germany. ⁹Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA. ¹⁰Section Clinical Tropical Medicine, Department of Infectious Diseases, Heidelberg University Hospital, INF 324, D 69120 Heidelberg, Germany. ¹¹Environmental Research Group Oxford (ERGO), Tinbergen Building, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. ¹²Oxford University Clinical Research Unit, Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam. ¹³Centre for Tropical Medicine, University of Oxford, Churchill Hospital, Oxford OX3 7LJ, UK. ¹⁴Department of Entomology, University of California Davis, Davis, California 95616, USA. ¹⁵Department of Medicine, National University of Singapore, 119228 Singapore.

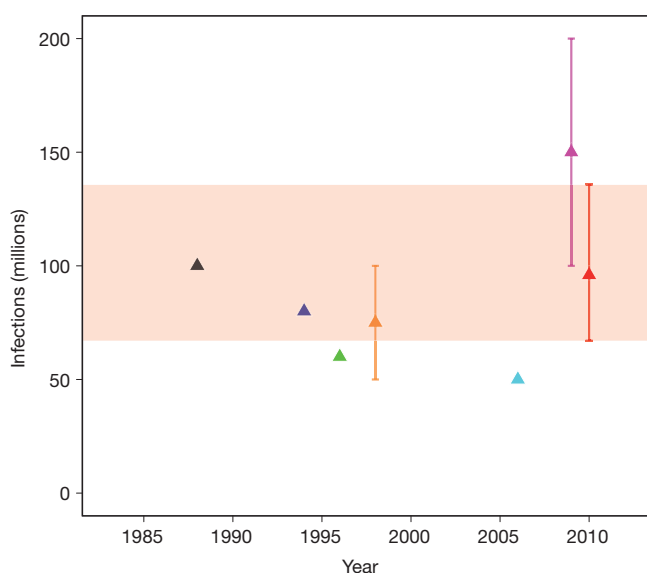


Figure 1 | Global estimates of total dengue infections. Comparison of previous estimates of total global dengue infections in individuals of all ages, 1985–2010. Black triangle, ref. 5; dark blue triangle, ref. 15; green triangle, ref. 17; orange triangle, ref. 16; light blue triangle, ref. 30; pink triangle, ref. 10; red triangle, apparent infections from this study. Estimates are aligned to the year of estimate and, if not stated, aligned to the publication date. Red shading marks the credible interval of our current estimate, for comparison. Error bars from ref. 10 and ref. 16 replicated the confidence intervals provided in these publications.

present/absent status⁴ (Fig. 2a); and built a boosted regression tree (BRT) statistical model of dengue risk that addressed the limitations of previous risk maps (see Supplementary Information, section C) to define the probability of occurrence of dengue infection (dengue risk) within each 5 km × 5 km pixel globally (Fig. 2b). The model was run 336 times to reflect parameter uncertainty and an ensemble mean map was created (see Supplementary Information, section C). We then combined this ensemble map with detailed longitudinal information on dengue infection incidence from cohort studies and built a non-parametric Bayesian hierarchical model to describe the relationship between dengue risk and incidence (see Supplementary Information, section D). Finally, we used the estimated relationship to predict the number of apparent and inapparent dengue infections in 2010 (see Supplementary Information, section E). Our definition of an apparent infection is consistent with that used by the cohort studies: an infection with sufficient severity to modify a person's regular schedule, such as attending school. This definition encompasses any level of severity of the disease.

We predict that dengue transmission is ubiquitous throughout the tropics, with the highest risk zones in the Americas and Asia (Fig. 2b). Validation statistics indicated high predictive performance of the BRT ensemble mean map with area under the receiver operating characteristic (AUC) of 0.81 (± 0.02 s.d., $n = 336$) (see Supplementary Information, section C). Predicted risk in Africa, although more unevenly distributed than in other tropical endemic regions, is much more widespread than suggested previously. Africa has the poorest record of occurrence data and, as such, increased information from this continent would help to define better the spatial distribution of dengue within it and to improve its derivative burden estimates. We found high levels of precipitation and temperature suitability for dengue transmission to be most strongly associated among the variables considered with elevated dengue risk, although low precipitation was not found to limit transmission strongly (see Supplementary Information, section C). Proximity to low-income urban and peri-urban centres was also linked to greater risk, particularly in highly connected areas, indicating that human movement between population centres is an important facilitator of dengue spread. These associations have previously

been cited⁹, but have not been demonstrated at the global scale and highlight the importance of including socioeconomic covariates when assessing dengue risk.

We estimate that there were 96 million apparent dengue infections globally in 2010 (Table 1). Asia bore 70% (67 (47–94) million infections) of this burden, and is characterized by large swathes of densely populated regions coinciding with very high suitability for disease transmission. India^{19,20} alone contributed 34% (33 (24–44) million infections) of the global total. The disproportionate infection burden borne by Asian countries is emphasized in the cartogram shown in Fig. 2c. The Americas contributed 14% (13 (9–18) million infections) of apparent infections worldwide, of which over half occurred in Brazil and Mexico. Our results indicate that Africa's dengue burden is nearly equivalent to that of the Americas (16 (11–22) million infections, or 16% of the global total), representing a significantly larger burden than previously estimated. This disparity supports the notion of a largely hidden African dengue burden, being masked by symptomatically similar illnesses, under-reporting and highly variable treatment-seeking behaviour^{6,9,20}. The countries of Oceania contributed less than 0.2% of global apparent infections.

We estimate that an additional 294 (217–392) million inapparent infections occurred worldwide in 2010. These mild or asymptomatic infections are not detected by the public health surveillance system and have no immediate implications for clinical management⁶. However, the presence of this huge potential reservoir of infection has profound implications for: (1) correctly enumerating economic impact (for example, how many vaccinations are needed to avert an apparent infection) and triangulating with independent assessments of disability adjusted life years (DALYs)²¹; (2) elucidating the population dynamics of dengue viruses²²; and (3) making hypotheses about population effects of future vaccine programmes²³ (volume, targeting efficacy, impacts in combination with vector control), which will need to be administered to maximize cross-protection and minimize post-vaccination susceptibility.

The absolute uncertainties in the national burden estimates are inevitably a function of population size, with the greatest uncertainties in India, Indonesia, Brazil and China (see full rankings in Supplementary Table 4). In addition, comparing the ratio of the mean to the width of the confidence interval²⁴ revealed the greatest contributors to relative uncertainty (see full rankings in Supplementary Table 4). These were countries with sparse occurrence points and low evidence consensus on dengue presence, such as Afghanistan or Rwanda (see Fig. 2a), or those with ubiquitous high risk, such as Singapore or Djibouti, for which our burden prediction confidence interval is at its widest (see Supplementary Information, section D, Fig. 2). Therefore, increasing evidence consensus and occurrence data availability in low consensus countries and assembling new cohort studies, particularly in areas of high transmission, will reduce uncertainty in future burden estimates. Our approach, uniquely, provides new evidence to help maximize the value and cost-effectiveness of surveillance efforts, by indicating where limited resources can be targeted to have their maximum possible impact in improving our knowledge of the global burden and distribution of dengue.

Our estimates of total infection burden (apparent and inapparent) are more than three times higher than the WHO predicted figure (Supplementary Information, section E). Our definition of an apparent infection is broad, encompassing any disruption to the daily routine of the infected individual, and consequently is an inclusive measurement of the total population affected adversely by the disease. Within this broad class, the severity of symptoms will affect treatment-seeking behaviours and the probability of a correct diagnosis in response to a given infection. Our definition is therefore more comprehensive than those of traditional surveillance systems which, even in the most efficient system, report a much narrower range of dengue infections. By reviewing our database of longitudinal cohort studies, in which total infections in the community were documented exhaustively, we find

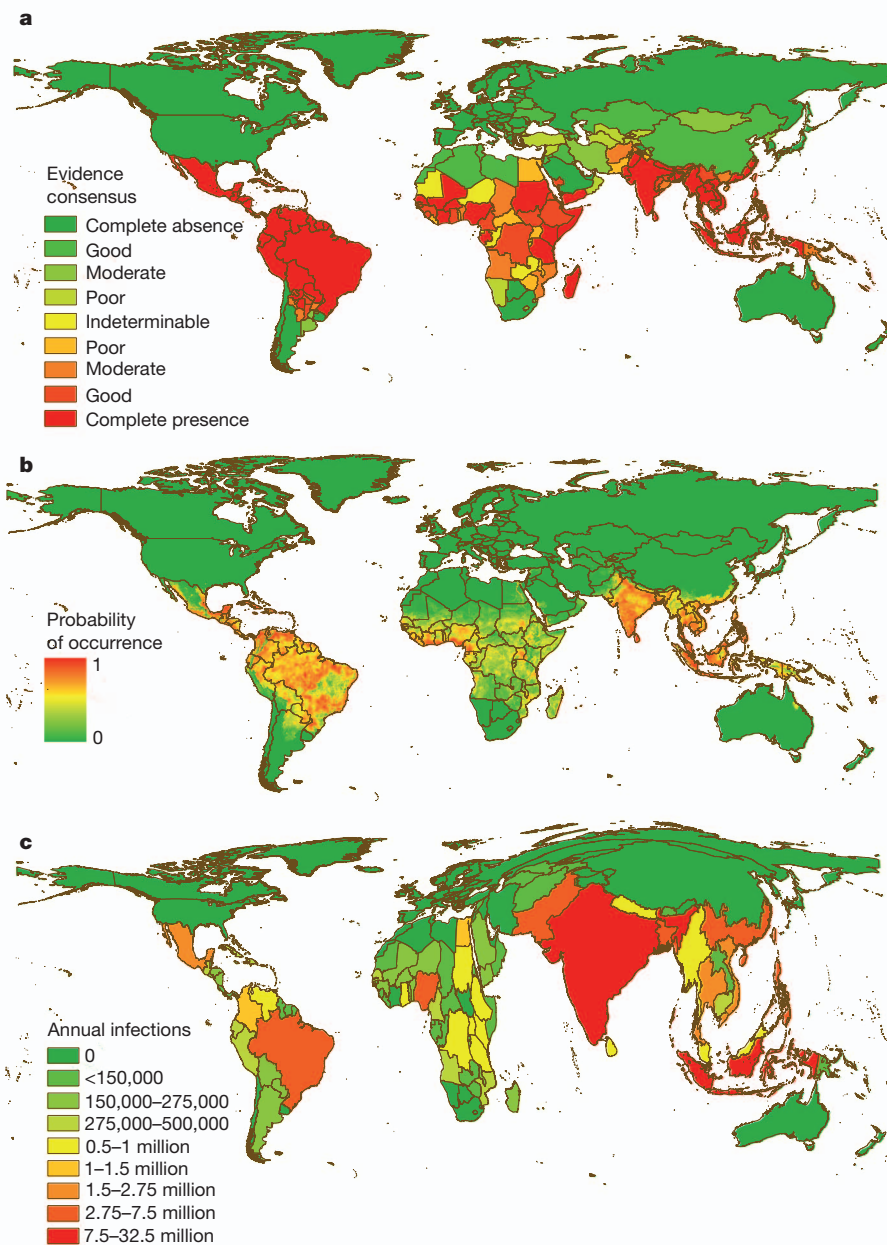


Figure 2 | Global evidence consensus, risk and burden of dengue in 2010. **a**, National and subnational evidence consensus on complete absence (green) through to complete presence (red) of dengue⁴. **b**, Probability of dengue occurrence at 5 km × 5 km spatial resolution of the mean predicted map (area under the receiver operator curve of 0.81 (±0.02 s.d., $n = 336$)) from 336

boosted regression tree models. Areas with a high probability of dengue occurrence are shown in red and areas with a low probability in green. **c**, Cartogram of the annual number of infections for all ages as a proportion of national or subnational (China) geographical area.

that the biggest source of disparity between actual and reported infection numbers is the low proportion of individuals with apparent infections seeking care from formal health facilities (see Supplementary Information, section E, Fig. 5 for full analysis). Additional biases are

introduced by misdiagnosis and the systematic failure of health management information systems to capture and report presenting dengue cases. By extracting the average magnitude of each of these sequential disparities from published cohort and clinical studies, we can recreate a hypothetical reporting chain with idealized reporting and arrive at estimates that are broadly comparable to those countries reported to the WHO. This is most clear in more reliable reporting regions such as the Americas. Systemic under-reporting and low hospitalization rates have important implications, for example, in the evaluation of vaccine efficacy based on reduced hospitalized caseloads. Inferences about these biases may be made from the comparison of estimated versus reported infection burdens in 2010, highlighting areas where particularly poor reporting might be strengthened (see Supplementary Information, section E).

Table 1 | Estimated burden of dengue in 2010, by continent

	Apparent	Inapparent
	Millions (credible interval)	Millions (credible interval)
Africa	15.7 (10.5–22.5)	48.4 (34.3–65.2)
Asia	66.8 (47.0–94.4)	204.4 (151.8–273.0)
Americas	13.3 (9.5–18.5)	40.5 (30.5–53.3)
Oceania	0.18 (0.11–0.28)	0.55 (0.35–0.82)
Global	96 (67.1–135.6)	293.9 (217.0–392.3)

We have strived to be exhaustive in the assembly of contemporary data on dengue occurrence and clinical incidence and have applied new modelling approaches to maximize the predictive power of these data. It remains the case, however, that the empirical evidence base for global dengue risk is more limited than that available, for example, for *Plasmodium falciparum*²⁵ and *Plasmodium vivax*²⁶ malaria. Records of disease occurrence carry less information than those of prevalence and, as databases of the latter become more widespread, future approaches should focus on assessing relationships between seroprevalence and clinical incidence as a means of assessing risk²⁷. Additional cartographic refinements are also required to help differentiate endemic from epidemic-prone areas, to determine the geographic diversity of dengue virus types and to predict the distributions of future risk under scenarios of socioeconomic and environmental change.

The global burden of dengue is formidable and represents a growing challenge to public health officials and policymakers. Success in tackling this growing global threat is, in part, contingent on strengthening the evidence base on which control planning decisions and their impact are evaluated. It is hoped that this evaluation of contemporary dengue risk distribution and burden will help to advance that goal.

METHODS SUMMARY

We compiled a database of 8,309 geo-located occurrence records for the period 1960 to 2012 from a combination of published literature and online resources¹⁸. All records were standardized annually (that is, repeat records in the same location within a year merged as one occurrence) and underwent rigorous quality control. From a suite of potential environmental and socioeconomic covariates, we chose a relevant subset including: (1) two precipitation variables interpolated from global meteorological stations; (2) an index of temperature suitability for dengue transmission adapted from an equivalent index for malaria²⁸; (3) a vegetation/moisture index; (4) demarcations of urban and peri-urban areas; (5) an urban accessibility metric; and (6) an indicator of relative poverty. We then built a disease distribution model using a boosted regression tree (BRT) framework. To compensate for the lack of absence data, we created an evidence-based probabilistic framework for generating pseudo-absences that mitigated the main biasing factors in pseudo-absence generation²⁹, namely: (1) geographical extent; (2) number; (3) contamination bias; and (4) sampling bias. We then created an ensemble of 336 BRT models using different plausible combinations of these factors and representing independent samples of possible sampling distributions. We calculated the final probability of occurrence (risk) map as the central tendency of these 336 BRT models predicted at a 5 km × 5 km resolution. Exclusion criteria were based on the definitive extents of dengue⁴ and temperature suitability for dengue transmission²⁸. Using detailed longitudinal information from 54 dengue cohort studies, we defined a relationship between the probability of dengue occurrence and inapparent and apparent incidence using a Bayesian hierarchical model. We defined a negative binomial likelihood function with constant dispersion and a rate characterized by a highly flexible data-driven Gaussian process prior. Uninformative hyperpriors were assigned hierarchically to the prior parameters and the full posterior distribution determined by Markov Chain Monte Carlo (MCMC) sampling. Using human population gridded data, estimates of dengue infections were then calculated nationally, regionally and globally for both apparent and inapparent infections.

Full Methods and any associated references are available in the online version of the paper.

Received 8 October 2012; accepted 7 March 2013.

Published online 7 April; corrected online 24 April 2013 (see full-text HTML version for details).

1. Simmons, C. P., Farrar, J. J., van Vinh Chau, N. & Wills, B. Dengue. *N. Engl. J. Med.* **366**, 1423–1432 (2012).
2. World Health Organization. *Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control*. WHO/HTM/NTD/DEN/2009.1 (World Health Organization, 2009).
3. Tatem, A. J., Hay, S. I. & Rogers, D. J. Global traffic and disease vector dispersal. *Proc. Natl Acad. Sci. USA* **103**, 6242–6247 (2006).
4. Brady, O. J. et al. Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS Negl. Trop. Dis.* **6**, e1760 (2012).
5. Halstead, S. B. Pathogenesis of dengue: challenges to molecular biology. *Science* **239**, 476–481 (1988).
6. Endy, T. P. et al. Determinants of inapparent and symptomatic dengue infection in a prospective study of primary school children in Kamphaeng Phet, Thailand. *PLoS Negl. Trop. Dis.* **5**, e975 (2011).

7. Sabchareon, A. et al. Protective efficacy of the recombinant, live-attenuated, CYD tetravalent dengue vaccine in Thai schoolchildren: a randomised, controlled phase 2b trial. *Lancet* **380**, 1559–1567 (2012).
8. Halstead, S. B. Dengue vaccine development: a 75% solution? *Lancet* **380**, 1535–1536 (2012).
9. Gubler, D. J. Dengue and dengue hemorrhagic fever. *Clin. Microbiol. Rev.* **11**, 480–496 (1998).
10. Beatty, M. E., Letson, G. W. & Margolis, H. S. Estimating the global burden of dengue. *Am. J. Trop. Med. Hyg.* **81** (Suppl. 1), 231 (2009).
11. Van Kleef, E., Bambrick, H. & Hales, S. The geographic distribution of dengue fever and the potential influence of global climate change. *Tropika.net* http://journal.tropika.net/scielo.php?script=sci_arttext&pid=S2078-86062010005000001&lng=en&nrm=iso (2009).
12. World Health Organization. *International Travel and Health: Situation as on 1 January 2012* (World Health Organization, 2012).
13. Hales, S., de Wet, N., Maindonald, J. & Woodward, A. Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. *Lancet* **360**, 830–834 (2002).
14. Rogers, D. J., Wilson, A. J., Hay, S. I. & Graham, A. J. The global distribution of yellow fever and dengue. *Adv. Parasitol.* **62**, 181–220 (2006).
15. Monath, T. P. Yellow fever and dengue: the interactions of virus, vector and host in the re-emergence of epidemic disease. *Semin. Virol.* **5**, 133–145 (1994).
16. Rigau-Pérez, J. G. et al. Dengue and dengue haemorrhagic fever. *Lancet* **352**, 971–977 (1998).
17. Rodhain, F. La situation de la dengue dans le monde. *Bull. Soc. Pathol. Exot.* **89**, 87–90 (1996).
18. Freifeld, C. C., Mandl, K. D., Reis, B. Y. & Brownstein, J. S. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *J. Am. Med. Inform. Assoc.* **15**, 150–157 (2008).
19. Chakravarti, A., Arora, R. & Luxemburger, C. Fifty years of dengue in India. *Trans. R. Soc. Trop. Med. Hyg.* **106**, 273–282 (2012).
20. Kakkar, M. Dengue fever is massively under-reported in India, hampering our response. *Br. Med. J.* **345**, e8574 (2012).
21. Murray, C. J. L. et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2197–2223 (2012).
22. Cummings, D. A. et al. The impact of the demographic transition on dengue in Thailand: insights from a statistical analysis and mathematical modeling. *PLoS Med.* **6**, e1000139 (2009).
23. Johansson, M. A., Hombach, J. & Cummings, D. A. Models of the impact of dengue vaccines: a review of current research and potential approaches. *Vaccine* **29**, 5860–5868 (2011).
24. Hay, S. I. et al. Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007. *PLoS Med.* **7**, e1000290 (2010).
25. Gething, P. W. et al. A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malar. J.* **10**, 378 (2011).
26. Gething, P. W. et al. A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. *PLoS Negl. Trop. Dis.* **6**, e1814 (2012).
27. Anders, K. L. & Hay, S. I. Lessons from malaria control to help meet the rising challenge of dengue. *Lancet Infect. Dis.* **12**, 977–984 (2012).
28. Gething, P. W. et al. Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*. *Parasites Vectors* **4**, 92 (2011).
29. Chefaoui, R. M. & Lobo, J. M. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecol. Modell.* **210**, 478–486 (2008).
30. TDR/World Health Organization. *Report of the Scientific Working Group on Dengue, 2006*. TDR/SWG/08 (TDR/World Health Organization, 2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements S.I.H. is funded by a Senior Research Fellowship from the Wellcome Trust (095066) which also supports S.B. and P.W.G. C.P.S. is also funded by a Senior Research Fellowship from the Wellcome Trust (084368). O.J.B. is funded by a BBSRC Industrial CASE studentship. J.P.M., A.W.F., T.J., G.R.W.W., C.P.S., T.W.S. and S.I.H. received funding from, and with S.B., P.W.G., O.J.B. and J.J.F. acknowledge the contribution of, the International Research Consortium on Dengue Risk Assessment Management and Surveillance (IDAMS, 21803, <http://www.idams.eu>). This work was funded in part by EU grant 2011-261504 EDENEXT and the paper is catalogued by the EDENEXT Steering Committee as EDENEXT. S.I.H. and T.W.S. also acknowledge funding support from the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health.

Author Contributions S.I.H. and J.J.F. conceived the research. S.B. and S.I.H. drafted the manuscript. S.B. drafted the Supplementary Information with significant support on sections A (O.J.B., C.L.M.), B (J.P.M., G.R.W.W.), C (P.W.G.), D (O.J.B., T.W.S.), and O.J.B. wrote section E. J.S.B. and A.G.H. provided HealthMap occurrence data and advice on its provenance. O.J.B. reviewed all the occurrence data. S.B. did the modelling and analysis with advice from J.M.D., P.W.G. and S.I.H. J.P.M. created all maps. All authors discussed the results and contributed to the revision of the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.I.H. (simon.hay@zoo.ox.ac.uk).

Diverse type VI secretion phospholipases are functionally plastic antibacterial effectors

Alistair B. Russell¹, Michele LeRoux^{1,2}, Krisztina Hathazi^{3,4}, Danielle M. Agnello¹, Takahiko Ishikawa^{3,4}, Paul A. Wiggins^{5,6}, Sun Nyunt Wai^{3,4} & Joseph D. Mougous^{1,2}

Membranes allow the compartmentalization of biochemical processes and are therefore fundamental to life. The conservation of the cellular membrane, combined with its accessibility to secreted proteins, has made it a common target of factors mediating antagonistic interactions between diverse organisms. Here we report the discovery of a diverse superfamily of bacterial phospholipase enzymes. Within this superfamily, we defined enzymes with phospholipase A₁ and A₂ activity, which are common in host-cell-targeting bacterial toxins and the venoms of certain insects and reptiles^{1,2}. However, we find that the fundamental role of the superfamily is to mediate antagonistic bacterial interactions as effectors of the type VI secretion system (T6SS) translocation apparatus; accordingly, we name these proteins type VI lipase effectors. Our analyses indicate that PldA of *Pseudomonas aeruginosa*, a eukaryotic-like phospholipase D³, is a member of the type VI lipase effector superfamily and the founding substrate of the haemolysin co-regulated protein secretion island II T6SS (H2-T6SS). Although previous studies have specifically implicated PldA and the H2-T6SS in pathogenesis^{3–5}, we uncovered a specific role for the effector and its secretory machinery in intra- and interspecies bacterial interactions. Furthermore, we find that this effector achieves its antibacterial activity by degrading phosphatidylethanolamine, the major component of bacterial membranes. The surprising finding that virulence-associated phospholipases can serve as specific antibacterial effectors suggests that interbacterial interactions are a relevant factor driving the continuing evolution of pathogenesis.

Within proteobacterial genomes, predicted lipases are often encoded adjacent to homologues of the *vgrG* gene^{3,6}. The VgrG protein is strongly associated with, and functionally important for, the cell-contact-dependent type VI secretion (T6S) protein delivery pathway⁷. This pathway, which is distributed throughout all classes of Proteobacteria, can target both eukaryotic and bacterial cells; however, it is the specificity of its effectors that dictates the consequences of intoxication by the system. Known T6S effectors are few and include enzymes that either modify actin or degrade peptidoglycan—both domain-restricted molecules^{8,9}. Thus, one would speculate that a barrier to the expansion or alteration of domain targeting would be the acquisition of a new effector or the evolution of one that is pre-existing.

To understand the significance of the T6S-associated lipases we undertook an informatic approach to examine their genetic context, sequence and phylogenetic distribution. This analysis uncovered 377 putative lipases comprising five divergent families (type VI lipase effectors 1–5 (Tle1–5)) that share no detectable overall sequence homology (Fig. 1a and Supplementary Figs 1–5). However, the families are united by a broad sporadic distribution pattern within Gram-negative bacteria and conserved putative catalytic motifs. Four of the families (Tle1–4) exhibit the GXSG motif common in esterases and many lipases, whereas the fifth (Tle5) possesses dual HXKXXXXD motifs found in phospholipase D (PLD) enzymes¹ (Fig. 1b). Outside

of catalytic motifs, Tle1–4 members lack significant homology with known lipase enzymes, suggesting that these proteins could represent previously uncharacterized diversity in the lipase superfamily.

Our previous work has shown that antibacterial T6S effectors are encoded adjacent to cognate immunity genes, which are essential owing to the self-targeting activity of the T6S apparatus^{9,10}. Moreover, because of a direct inactivation mechanism, the localization of the immunity protein indicates the cellular compartment targeted by the effector. Examination of the genomic context of the putative lipase-encoding genes revealed each is found adjacent to an open reading frame encoding a predicted periplasmic protein (Fig. 1a). Thus, we propose that contrary to prevailing views of bacterial lipase function, *vgrG*-associated lipase families could universally serve roles in interbacterial competition, possibly targeting phospholipids accessible from the periplasm. Consistent with our hypothesis, one of the putative lipase enzymes that we identified, *Vibrio cholerae* VC1418 (Tle2^{VC}; Supplementary Fig. 2), was recently found to act as an effector in amoeba defence and intra-species bacterial competition¹¹. Although the biochemical activity of *V. cholerae* Tle2 was not determined, this suggests a capacity for Tle proteins to target a structure conserved in eukaryotes and bacteria.

To determine whether Tle2^{VC} participates in interspecies bacterial antagonism, we tested its ability to provide fitness to *V. cholerae* in competition with *Escherichia coli*. We observed that *V. cholerae* strains lacking *tle2*^{VC} display a marked impairment in their capacity to kill *E. coli*, approaching that of a strain lacking T6S function (Fig. 2a and Supplementary Fig. 6). It is of note that a previous study probing the function of Tle2^{VC} did not observe a contribution of the protein to fitness in an interspecies setting¹¹. This study was performed with strain V52, in which T6S-associated genes exhibit constitutively high expression¹². Therefore, a potential explanation for the apparent discrepancy is that in a hyperactive state the absence of one effector is not sufficient to diminish antibacterial activity to a measurable level.

Tle2 represents only one of four divergent GXSG families within the broader superfamily. As a first step towards understanding the functional importance of other GXSG families, we examined *Burkholderia thailandensis* BTH_I2698 (Tle1^{BT}; Fig. 1 and Supplementary Fig. 1), which we previously demonstrated to be a substrate of an antibacterial T6SS (ref. 10). The *tle1*^{BT} gene is found adjacent to genes encoding two homologous periplasmic lipoproteins, I2699 and I2700, which we posited could serve as Tle1 immunity proteins. Furthermore, *tle1*^{BT}, I2699 and I2700 seem to have been subject to a duplication event, with homologues of all three genes present immediately upstream (Supplementary Fig. 7). To simplify our analysis, we generated a mutant strain lacking one copy of this duplicated region. Using labelled derivatives of this strain co-cultured under T6S-conductive conditions, we found that recipient strains lacking *tle1*^{BT} and its putative immunity determinants exhibit significantly decreased fitness in competition with donor strains possessing *tle1*^{BT} and a functional T6SS, and that expression of I2699 in the recipient strain was necessary

¹Department of Microbiology, University of Washington, Seattle, Washington 98195, USA. ²Molecular and Cellular Biology Program, University of Washington, Seattle, Washington 98195, USA.

³Department of Molecular Biology, Umeå University, SE-90187 Umeå, Sweden. ⁴The Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, SE-90187 Umeå, Sweden. ⁵Department of Physics, University of Washington, Seattle, Washington 98195, USA. ⁶Department of Bioengineering, University of Washington, Seattle, Washington 98195, USA.

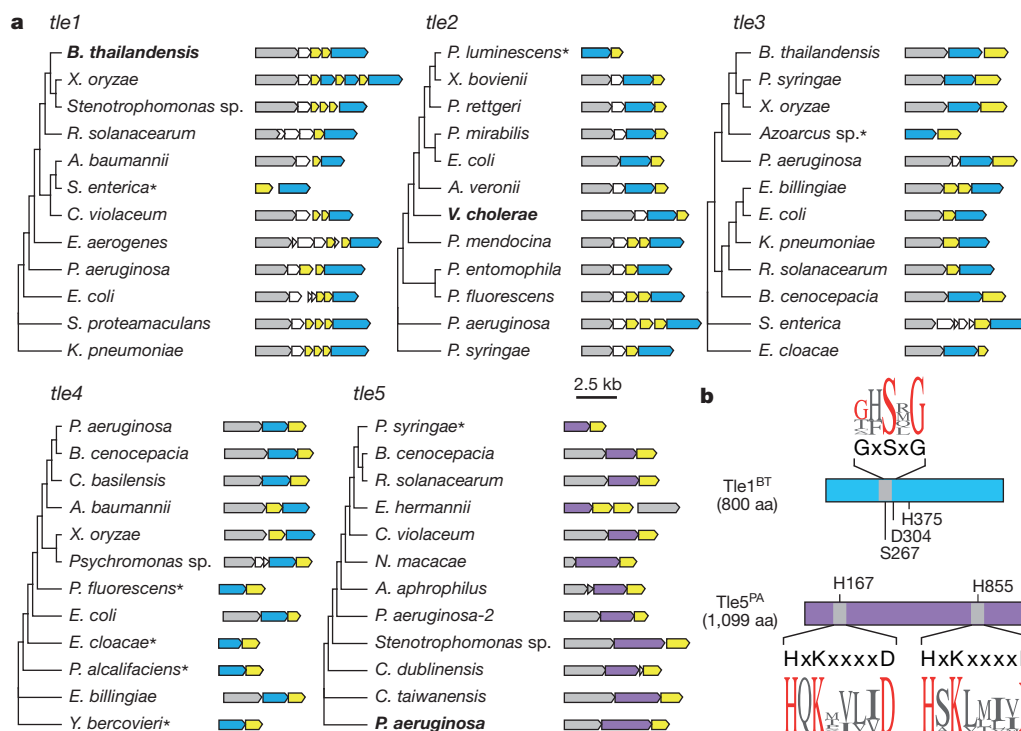


Figure 1 | Overview of the Tle superfamily. **a**, Evolutionary trees, genetic organization, and phylogenetic distribution of select Tle family members. Genes are coloured by their predicted protein product (blue, Tle proteins with a GXSG catalytic motif; purple, Tle proteins with dual HXKXXXXD catalytic motifs; grey, VgrG proteins; yellow, putative periplasmic immunity proteins). Branch lengths are not proportional to evolutionary distance. Asterisks denote

and sufficient to restore competitive fitness (Fig. 2b). These data show that Tle1^{BT} is an antibacterial effector delivered between cells by T6S, and that I2699, hereafter referred to as type VI secretion lipase immunity 1 (Tli1^{BT}), protects against Tle1^{BT}.

Having demonstrated that members of two GXSG Tle families function as antibacterial T6S effectors, we next sought to investigate their biochemical activity. To characterize Tle1^{BT} and Tle2^{VC}, we purified the proteins and catalytic nucleophile substitution mutant derivatives (Tle1^{BT}(Ser267Ala) and Tle2^{VC}(Ser371Ala)) as amino-terminal fusions to hexahistidine (His₆)-tagged maltose binding protein (His₆-MBP), which we found to be necessary to generate and maintain soluble protein (Supplementary Figs 8 and 9). Notably, Tle1–4 possess a Ser-Asp-His catalytic triad used by a diversity of esterase enzymes, including thioesterases, acylesterases and assorted lipase and phospholipases¹. Given this wide range of potential activities, we initially confirmed general esterase activity of His₆-MBP–Tle1^{BT} and His₆-MBP–Tle2^{VC} by demonstrating that these effectors, but not their catalytic substitution mutants, hydrolyse a model substrate, polysorbate 20 (Supplementary Fig. 10). Next we asked whether His₆-MBP–Tle1^{BT} or His₆-MBP–Tle2^{VC} possesses phospholipase activity. Using vesicle substrates doped with fluorescent phospholipid derivatives, we determined that His₆-MBP–Tle1^{BT} acts specifically as a phospholipase A₂ (PLA₂), and His₆-MBP–Tle2^{VC} as a phospholipase A₁ (PLA₁) (Fig. 2c, d). Linking these activities to the antibacterial phenotypes we observed associated with the proteins *in vivo*, neither Tle1^{BT}(Ser267Ala) nor Tle2^{VC}(Ser371Ala), both catalytically inactive, serve as antibacterial effectors (Fig. 2a, b). Moreover, we found that the PLA₂ activity of His₆-MBP–Tle1^{BT} is robustly inhibited by the addition of its immunity protein, Tli1^{BT} (Fig. 2e).

If GXSG family Tle proteins serve as antibacterial T6SS phospholipases, we reasoned that their activity against sensitive recipients should correlate with an increase in cellular permeability. To test these predictions, we performed single-cell measurements of propidium

tle genes without an apparent adjacent *vgrG* gene. **b**, Domain organization of a single member of the GXSG and dual HXKXXXXD catalytic classes of Tle proteins. Regions comprising these catalytic motifs are labelled in grey, and positions of all putative catalytic residues are denoted. Sequence logos were generated from alignments of the catalytic motifs from Tle1–4 (GXSG) and catalytic motifs from Tle5 (HXKXXXXD). aa, amino acids; kb, kilobases.

iodide uptake within interbacterial competitions of *B. thailandensis*. Consistent with our hypotheses, the lack of Tle1 immunity within cells corresponded to significantly increased propidium iodide uptake (Fig. 2f and Supplementary Videos 1–3). Using automated cell identity and tracking algorithms¹³, we further demonstrated that the increase in propidium iodide uptake depended on direct contact with donor cells possessing a functional T6SS (Fig. 2g and Supplementary Fig. 11).

With our data validating members of two GXSG families as antibacterial phospholipase effectors, we explored whether these findings could be extended to the HXKXXXXD family (Tle5). This catalytic motif is strongly indicative of PLD activity¹, which has previously not been associated with an antibacterial enzyme. We choose *Pseudomonas aeruginosa* PldA, hereafter referred to as Tle5^{PA} (Fig. 1 and Supplementary Fig. 5), as a representative Tle5 family member. We began our study by confirming the enzymatic activity of the protein, as its function was previously studied in the context of cellular extracts³. Consistent with previous observations, Tle5^{PA} catalyses the release of choline from phosphatidylcholine, in a manner dependent on a predicted catalytic histidine residue (His 855) (Fig. 3a and Supplementary Fig. 12). Under similar conditions neither Tle1^{BT} nor Tle2^{VC} showed appreciable activity in this assay, underscoring the diverse substrate specificity within the Tle superfamily.

A candidate Tle5^{PA} periplasmic immunity protein is not readily apparent, as the adjacent gene, PA3488, is predicted to encode a cytoplasmic protein. However, expression of PA3488 from a second, upstream, predicted start site yields a periplasmically localized protein, hereafter referred to as Tli5^{PA}, that binds specifically to Tle5^{PA} (Supplementary Fig. 13). To probe the role of Tle5^{PA} and Tli5^{PA} in interbacterial interactions, we generated a lysis reporter strain bearing a deletion of the *tle5*^{PA} *tli5*^{PA} bicistron. Lysis of this strain was greatly increased when co-cultured with a wild-type, but not a Δ *tle5*^{PA} donor, strain (Fig. 3b). Furthermore, expression of *tli5*^{PA} in the recipient was sufficient to protect from Tle5^{PA}-dependent lysis. Together, these data

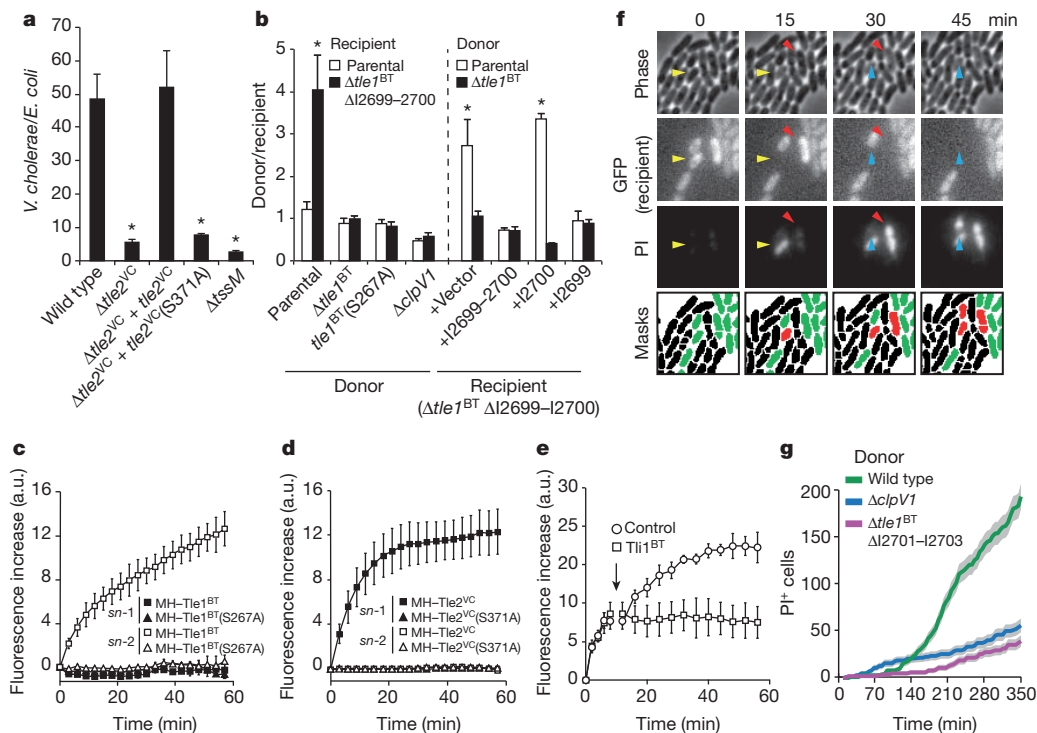


Figure 2 | Tle GXSXG-type proteins are antibacterial phospholipase effectors delivered by the T6SS. **a**, Outcome of growth competitions between the indicated *V. cholerae* strains and *E. coli*. The $\Delta tssM$ strain is inactive for T6SS. Asterisks denote competitive outcomes significantly different than those obtained with wild type ($P < 0.05$, $n = 3$). **b**, Growth competition assays between the indicated *B. thailandensis* donor and recipient strains. The $\Delta clpV1$ strain is inactivated for T6SS-1, which is required for Tle1^{BT} export¹⁰. The parental strain is $\Delta I2701-2703$. Asterisks denote competition outcomes significantly different between indicated recipient strains (left) or indicated donor strains (right) ($P < 0.05$, $n = 3$). **c**, **d**, Enzymatic activity of the designated proteins against vesicles containing phospholipid derivatives with fluorescent moieties at the stereochemical numbering (*sn*) positions *sn*-1 and *sn*-2 ($n = 4$). a.u., arbitrary units; MH, His₆-MBP. **e**, Enzymatic activity of His₆-MBP-Tle1^{BT}

demonstrate that Tle5^{PA} acts as an antibacterial toxin and that Tli5^{PA} is its cognate immunity determinant.

The *P. aeruginosa* genome encodes three T6SSs, the H1–3-T6SSs. The H1-T6SS is the only system with known substrates and a demonstrated role in interbacterial interactions¹⁴. To define the T6SS involved in Tle5^{PA} transport, we constructed strains bearing individual in-frame deletions of the crucial ATPase genes, *clpV1*–3, associated with the H1–3 systems, respectively. Specific inactivation of the H2-T6SS in a donor strain abrogated Tle5^{PA}-dependent toxicity, indicating that this system is responsible for Tle5^{PA} delivery (Fig. 3b).

The finding that Tle5^{PA} transits the H2-T6S pathway is interesting in light of data that implicate this T6SS as a virulence factor in plant, mammalian cell culture, worm, and mouse models of infection^{4,5}. To more thoroughly explore the role of Tle5^{PA} and the H2-T6SS in interbacterial interactions, we measured their influence on competition outcomes between *P. aeruginosa* and a model T6S target, *Pseudomonas putida*⁹. Our results showed that both Tle5^{PA} and the H2-T6SS significantly contribute to the fitness of *P. aeruginosa* in interspecies competition under T6S-conductive conditions (Fig. 3c and Supplementary Fig. 14). These findings show that Tle5^{PA} is a potent antibacterial effector delivered by the H2-T6SS.

Although our data thus far show that Tle1^{BT}, Tle2^{VC} and Tle5^{PA} possess phospholipase activity *in vitro*, this did not allow us to assign definitively the toxic consequences of these effectors to membrane destruction. The phospholipase activity of the effectors could be accessory to a second toxicity mechanism found in these large, multi-domain proteins. To resolve this remaining ambiguity concerning

on *sn*-2-labelled phospholipids as measured in **c** after the addition of the indicated immunoprecipitate (arrow) ($n = 5$). **f**, Representative cropped micrograph series displaying three propidium iodide (PI) uptake and subsequent lysis events in a growth competition experiment between wild-type *B. thailandensis* and a Tle1^{BT}-sensitive recipient, $\Delta I2698-I2703$. Each event spans two frames and is highlighted by arrowheads. The mask frames depict cell assignments made by gating cells based on fluorescence (black, donor; green, recipient; red, propidium iodide-positive). Original magnification, $\times 60$. GFP, green fluorescent protein. **g**, Quantification of propidium iodide staining events from *B. thailandensis* growth competitions using automated custom software. The recipient strain was the same as in **f**. Shading indicates counting error. All error bars denote s.d.

Tle function, we focused our studies on Tle5^{PA}. Because a mixture of healthy and intoxicated cells could complicate our measurements, we decided to assay Tle5^{PA} effects in self-intoxicating monocultures of $\Delta tli5^{PA}$, in which each cell serves both as a donor and a sensitive recipient. As expected, this strain exhibited increased membrane permeability in a manner dependent on an active H2-T6SS and Tle5^{PA} (Supplementary Fig. 15).

Under conditions promoting intercellular delivery of Tle5^{PA}, we extracted lipids from both non-intoxicated (wild type) and intoxicated ($\Delta tli5$) cells, and quantified their phospholipid composition using mass spectrometry. This analysis showed that the unchecked action of Tle5^{PA} leads to a severely perturbed membrane phospholipid composition. Notably, phosphatidic acid, a product of PLD activity and a minor constituent of wild-type membranes (0.17%), was present at 8.1% in $\Delta tli5^{PA}$ —a 48-fold enrichment (Fig. 3d and Supplementary Table 1). The increased phosphatidic acid seemed to derive primarily from phosphatidylethanolamine, as it underwent a concomitant decrease of similar magnitude. Finally, we noted that phosphatidylglycerol increased slightly in $\Delta tli5$ relative to the wild type. We speculate that this latter result derives either from a compensatory effect or from Tle5^{PA} activity against cardiolipin, a minor component of *P. aeruginosa* membranes not detectable by the analysis method we used. Taken together, these data strongly suggest that Tle5^{PA}-imposed cell death occurs through phosphatidic acid accumulation via PLD activity, primarily directed against phosphatidylethanolamine. The precise physiological consequences of massive phosphatidic acid accumulation in bacterial cells are not known; however, the strong negatively

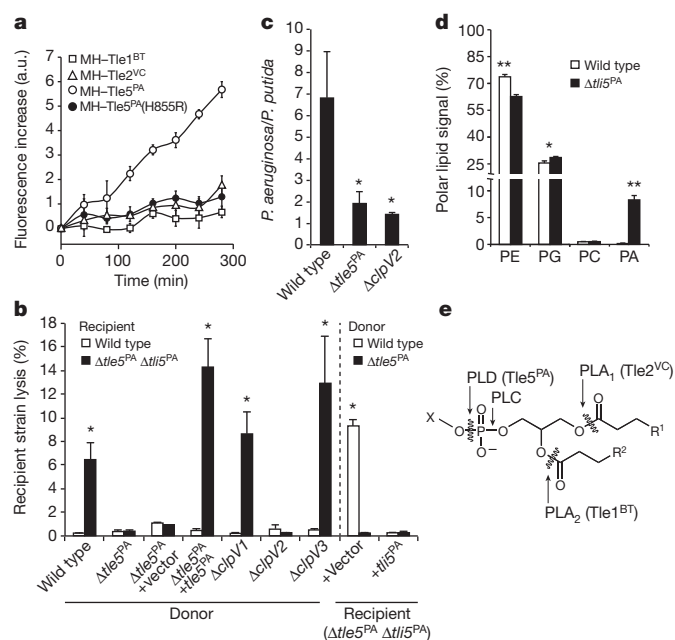


Figure 3 | Tle5^{PA} is an HXXXXXXD-type interspecies antibacterial phospholipase effector delivered by the H2-T6SS of *P. aeruginosa*.

a, Phosphatidylcholine-specific PLD activity of the indicated proteins against mixed lipid vesicles ($n = 3$). **b**, Lysis of recipient strains grown in co-culture with the indicated donor strains. Asterisks mark experiments in which recipient lysis is significantly different between indicated recipients (left), or between indicated donors (right) ($P < 0.05$, $n = 3$). **c**, Competitive growth of indicated *P. aeruginosa* strains against *P. putida* under T6SS-conductive conditions. Asterisks denote competition outcomes significantly different than those obtained with wild-type *P. aeruginosa* ($P < 0.05$, $n = 3$). **d**, Summary of phospholipid profiles of the indicated *P. aeruginosa* strains. Statistical significance noted ($n = 4$, $*P < 0.01$, $**P < 0.001$). PA, phosphatidic acid; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PG, phosphatidylglycerol. **e**, Generalized schematic of a phospholipid indicating the activities defined in this study. All error bars denote s.d.

charged character of the molecule is likely to have a detrimental effect on both integral and peripheral membrane-associated proteins. It is known that phosphatidic acid induces membrane curvature that can promote fusion and fission events¹⁵; therefore, Tle5^{PA} activity might also lead to generalized membrane destabilization, membrane blebbing and depolarization. Interestingly, the *in vivo* specificity of Tle5^{PA} for phosphatidylethanolamine, the major phospholipid constituent of most bacterial membranes, affords *P. aeruginosa* the capacity to use this enzyme against a vast array of competitors.

The discovery of T6SS-delivered phospholipase effectors has many implications. Crucially, their biochemical activity does not intrinsically limit their toxicity to bacterial cells (Fig. 3e). Indeed, two specificities now ascribed to Tle superfamily members, PLD and PLA₂, are both highly represented in host cell-targeting bacterial toxins². As these effectors are found in numerous established and emerging opportunistic pathogens, our work highlights the need to understand the biochemical, genetic and evolutionary basis of interdomain targeting by the T6SS. Such knowledge may ultimately become a component of a larger strategy to develop predictive algorithms for the evolution of bacterial pathogens. In addition, our findings add a new dimension to our understanding of the mechanisms used during bacterial competition. On the basis of our data it appears that membrane targeting evolved independently on several occasions as an antibacterial strategy. This convergent evolution underscores the susceptibility of the bacterial membrane to attack, a theme mirrored by the previous observation that bacteriolytic T6S effectors likewise degrade an essential, conserved bacterial structure⁹. The continued discovery of antibacterial effectors promises to illuminate further vulnerabilities of the

bacterial cell, and thus may aid our efforts to define promising therapeutic targets.

METHODS SUMMARY

B. thailandensis, *V. cholerae* and *P. aeruginosa* strains used in this study were derived from the strains E264, A1552 and PAO1, respectively^{16–18}. All deletions were in-frame, unmarked, and generated by allelic exchange. All protein and nucleotide sequences were obtained from the NCBI GenBank database. Informatic analysis of Tle protein identity, distribution, alignment and phylogeny used a combination of structural prediction, homology, subcellular localization and tree-building algorithms as detailed in the Methods. Tle proteins were purified as fusions to both MBP and a His₆ tag. Lipase activity was measured using fluorescent reporters as detailed in the Methods. Bacterial competition experiments were performed on Luria–Bertani (*V. cholerae* and *B. thailandensis*) or synthetic cystic fibrosis sputum media (SCFM)¹⁹ agar as detailed in the Methods. *P. aeruginosa* lysis assays were performed as previously described on competitions grown at 23 °C on 1.5% (w/v) agar SCFM plates²⁰. *P. aeruginosa* propidium iodide staining experiments were performed on monocultures grown at 23 °C on 1.5% (w/v) agar SCFM plates. Real-time single-cell quantification of propidium iodide staining under competition conditions used a modification of previously described custom software for the analysis of labelled bacteria visualized by fluorescence microscopy¹³. Lipidomic studies were performed by the Kansas State University Lipidomics Research Center.

Full Methods and any associated references are available in the online version of the paper.

Received 16 January; accepted 15 March 2013.

Published online 3 April 2013.

- Aloulou, A., Ali, Y. B., Bezzine, S., Gargouri, Y. & Gelb, M. H. Phospholipases: an overview. *Methods Mol. Biol.* **861**, 63–85 (2012).
- Schmiel, D. H. & Miller, V. L. Bacterial phospholipases and pathogenesis. *Microbes Infect.* **1**, 1103–1112 (1999).
- Wilderman, P. J., Vasil, A. I., Johnson, Z. & Vasil, M. L. Genetic and biochemical analyses of a eukaryotic-like phospholipase D of *Pseudomonas aeruginosa* suggest horizontal acquisition and a role for persistence in a chronic pulmonary infection model. *Mol. Microbiol.* **39**, 291–304 (2001).
- Lesic, B., Starkey, M., He, J., Hazan, R. & Rahme, L. G. Quorum sensing differentially regulates *Pseudomonas aeruginosa* type VI secretion locus I and homologous loci II and III, which are required for pathogenesis. *Microbiology* **155**, 2845–2855 (2009).
- Sana, T. G. et al. The second type VI secretion system of *Pseudomonas aeruginosa* strain PAO1 is regulated by quorum sensing and Fur and modulates internalization in epithelial cells. *J. Biol. Chem.* **287**, 27095–27105 (2012).
- Barret, M., Egan, F., Fargier, E., Morrissey, J. P. & O’Gara, F. Genomic analysis of the type VI secretion systems in *Pseudomonas* spp.: novel clusters and putative effectors uncovered. *Microbiology* **157**, 1726–1739 (2011).
- Silverman, J. M., Brunet, Y. R., Cascales, E. & Mougous, J. D. Structure and regulation of the type VI secretion system. *Annu. Rev. Microbiol.* **66**, 453–472 (2012).
- Pukatzki, S., Ma, A. T., Revel, A. T., Sturtevant, D. & Mekalanos, J. J. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc. Natl Acad. Sci. USA* **104**, 15508–15513 (2007).
- Russell, A. B. et al. Type VI secretion delivers bacteriolytic effectors to target cells. *Nature* **475**, 343–347 (2011).
- Russell, A. B. et al. A widespread bacterial type VI secretion effector superfamily identified using a heuristic approach. *Cell Host Microbe* **11**, 538–549 (2012).
- Dong, T. G., Ho, B. T., Yoder-Himes, D. R. & Mekalanos, J. J. Identification of T6SS-dependent effector and immunity proteins by Tn-seq in *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA* **110**, 2623–2628 (2013).
- Ishikawa, T. et al. Pathoadaptive conditional regulation of the type VI secretion system in *Vibrio cholerae* O1 strains. *Infect. Immun.* **80**, 575–584 (2012).
- LeRoux, M. et al. Quantitative single-cell characterization of bacterial interactions reveals type VI secretion is a double-edged sword. *Proc. Natl Acad. Sci. USA* **109**, 19804–19809 (2012).
- Mougous, J. D. et al. A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science* **312**, 1526–1530 (2006).
- Stace, C. L. & Ktistakis, N. T. Phosphatidic acid- and phosphatidylserine-binding proteins. *Biochim. Biophys. Acta* **1761**, 913–926 (2006).
- Kim, H. S. et al. Bacterial genome adaptation to niches: divergence of the potential virulence genes in three *Burkholderia* species of different survival strategies. *BMC Genomics* **6**, 174 (2005).
- Yildiz, F. H. & Schoolnik, G. K. Role of *rpoS* in stress survival and virulence of *Vibrio cholerae*. *J. Bacteriol.* **180**, 773–784 (1998).
- Stover, C. K. et al. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**, 959–964 (2000).
- Palmer, K. L., Aye, L. M. & Whiteley, M. Nutritional cues control *Pseudomonas aeruginosa* multicellular behavior in cystic fibrosis sputum. *J. Bacteriol.* **189**, 8079–8087 (2007).

20. Chou, S. *et al.* Structure of a peptidoglycan amidase effector targeted to Gram-negative bacteria by the type VI secretion system. *Cell Rep.* **1**, 656–664 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank H. Kulasekara, B. E. Uhlin and members of the Mougous and Wai laboratories for discussions, J. Woodward for sharing chemistry expertise, the Manoil laboratory for sharing *B. thailandensis* transposon mutants, and the Parsek laboratory and K. Korotkov for sharing reagents. This work was supported by grants from the National Institutes of Health (NIH) (AI080609, AI057141 and AI105268), Cystic Fibrosis Foundation (CFR565-CR07), the National Science Foundation (PHY-084845 and MCB-1151043), the Swedish Research Council (2010-3073, 2007-8673 UCMR Linnaeus and 2006-7431 MIMS), and the Faculty of Medicine, Umeå University. A.B.R. was supported by a Graduate Research Fellowship from the National Science Foundation (DGE-0718124), M.L. was supported by the NIH Cellular

and Molecular Training Grant (GM07270), P.A.W. received support from the University of Washington Royalty Research Fund and the Sloan Foundation, and J.D.M. holds an Investigator in the Pathogenesis of Infectious Disease Award from the Burroughs Wellcome Fund.

Author Contributions A.B.R., M.L., P.A.W., S.N.W. and J.D.M. conceived and designed experiments. A.B.R., M.L., K.H., D.M.A., T.I. and J.D.M. conducted experiments. A.B.R. and J.D.M. wrote the paper.

Author Information GenBank accession numbers for all Tle proteins identified in this study are found in Supplementary Figs 1–5. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.D.M. (mougous@u.washington.edu).

METHODS

Bacterial strains and growth conditions. *B. thailandensis* strains used in this study were derived from the sequenced strain E264 (ref. 16). *B. thailandensis* strains were grown on either Luria-Bertani (LB) medium, or the equivalent lacking additional NaCl (low-salt LB: 10 g bacto-peptone and 5 g yeast extract per litre) at 37 °C, supplemented with 200 µg ml⁻¹ trimethoprim and 25 µg ml⁻¹ irgasan where necessary. *V. cholerae* strains used in this study were derived from the O1 El Tor strain A1552 (ref. 17). *V. cholerae* was grown on LB medium or LB with 340 mM NaCl at 37 °C or 30 °C, supplemented with 100 µg ml⁻¹ rifampin, 100 µg ml⁻¹ carbenicillin and stated concentrations of arabinose as needed. *P. aeruginosa* strains used in this study were derived from the sequenced strain PAO1 (ref. 18). *P. aeruginosa* strains were grown on LB medium at 37 °C or synthetic cystic fibrosis sputum media (SCFM)¹⁹ at 23 °C supplemented with 25 µg ml⁻¹ irgasan, 30 µg ml⁻¹ gentamycin, and stated concentrations of isopropyl-β-D-thiogalactoside (IPTG) as required. *P. aeruginosa* lysis reporter strains were generated as described previously²⁰. For introducing in-frame deletions, *B. thailandensis* was grown on M9 minimal medium agar plates with 0.4% glucose as a carbon source and 0.1% (w/v) *p*-chlorophenylalanine for counter-selection²¹. *V. cholerae* was grown on LB supplemented with 10% (w/v) sucrose at 30 °C for counter-selection^{22,23}, and *P. aeruginosa* was grown on low-salt LB supplemented with 5% (w/v) sucrose at 30 °C for counter-selection²⁴. *P. putida* used in this study was the sequenced strain KT2440 (ref. 25). *P. putida* was grown on LB medium at 30 °C or on SCFM agar at 23 °C. *E. coli* strains used in this study included DH5α for plasmid maintenance and production of Tli1^{BT} immunoprecipitate, SM10 λpir for conjugal transfer of plasmids into *B. thailandensis*, *V. cholerae* and *P. aeruginosa*, MC4100 for competition assays with *V. cholerae*, BL21 (DE3) plyS for Tle5^{PA} immunoprecipitation studies, and Shuffle T7 plyS Express (New England Biolabs) for purification of Tle proteins. All *E. coli* strains were grown on LB medium or 2× yeast extract and tryptone (2× YT) medium at 37 °C, supplemented with 150 µg ml⁻¹ carbenicillin, 50 µg ml⁻¹ kanamycin, 30 µg ml⁻¹ chloramphenicol, 200 µg ml⁻¹ trimethoprim, 50 µg ml⁻¹ streptomycin, 15 µg ml⁻¹ gentamycin, 0.1% rhamnose and 100 mM IPTG as needed.

DNA manipulations. The creation, maintenance and transformation of plasmid constructs followed standard molecular cloning procedures. All primers used in this study were obtained from Integrated DNA Technologies. DNA amplification was carried out using either Phusion (New England Biolabs) or Mangomix (BioLine). DNA sequencing was performed by Genewiz Incorporated. Restriction enzymes were obtained from New England Biolabs. Splice overlap excision (SOE) PCR was performed as previously described²⁶.

Plasmid construction. Plasmids used for expression in this study were pET28b:His₆-MBP-TEV-His₆ (ref. 27), pET22b+ (Novagen) and pScrhaB2 (ref. 28) for *E. coli*, pPSV35CV (ref. 29) for *P. aeruginosa*, and pBAD24 (ref. 30) for *V. cholerae*. Complementation in *B. thailandensis* was performed using the Tn7-based integration vector pUC18T-miniTn7T-Tp::PS12 (ref. 31). In-frame deletions were generated using the suicide vectors pJRC115 for *B. thailandensis*²¹, pVCD442 for *V. cholerae*³², and pEXG2 for *P. aeruginosa*²⁴. For the production of deletion constructs, either 600-base-pair (bp) (*B. thailandensis* and *P. aeruginosa*) or 500-bp (*V. cholerae*) regions flanking the deletion were amplified, ligated together using SOE PCR, and subsequently cloned into pJRC115, pEXG2 or pVCD442, respectively. To generate the Tle1^{BT} (Ser267Ala) *B. thailandensis* mutation construct, 600-bp regions flanking the mutation with an additional overlapping extension consisting of the desired mutation were amplified and ligated together using SOE PCR and subsequently cloned into pJRC115. For *B. thailandensis* complementation constructs genes were amplified along with predicted ribosomal-binding sites and cloned into pUC18T-miniTn7T-Tp::P12. For *P. aeruginosa* complementation and expression constructs, genes were amplified with their native ribosomal-binding sites into pPSV35CV with a 3' fusion to the vesicular stomatitis virus glycoprotein (VSV-G) epitope tag. To generate the pPSV35CV::tle5^{PA}(H167R) and (H855R)-V constructs, the entire *tle5*^{PA} gene was amplified from pPSV35CV::tle5^{PA} and SOE PCR was used to introduce the desired base pair mutations. For pScrhaB2 *E. coli* expression constructs and pBAD24 *V. cholerae* expression and complementation constructs, genes were cloned downstream of the optimized ribosomal-binding site already present in these vectors with a fusion to a 3' VSV-G linker. To generate the pBAD24::tle2^{VC}(Ser371Ala)-V construct, the entire *tle2*^{VC} gene was amplified from pBAD24::tle2^{VC}-V and SOE PCR was used to introduce the desired base-pair mutations. This product was subsequently cloned into pBAD24. For the Tle purification constructs, *tle* genes were amplified and cloned into pET28b:His₆-MBP-TEV-His₆ to generate an N-terminal fusion to an MBP protein and a His₆ purification tag. SOE PCR was then used to generate the desired catalytic nucleophile substitution mutants. For the Tle5^{PA} periplasmic expression construct, *tle5*^{PA} was amplified and cloned into pET22b+ to generate an N-terminal fusion to the PelB leader peptide and a carboxy-terminal fusion to a His₆-epitope tag.

Informatic identification of Tle proteins. All sequences were obtained from the NCBI, and GenBank accession numbers for all Tle proteins identified in this study are found in Supplementary Figs 1–5. BTH_12698 from *B. thailandensis* E264, PA0260, PA1510 and PA3487, and PA5089 from *P. aeruginosa* PAO1, and VC1418 from *V. cholerae* V52, all encoded adjacent to *vgrG* genes, were identified as putative lipases using the PHYRE 2 structural prediction server³³. Using the amino acid sequences of these predicted lipases, blastp analyses were performed against the non-redundant protein database (<http://ftp.ncbi.nih.gov/blast/db/>) to identify unique instances of their homologues. Homology identified by the blast server was used to distribute these proteins into five distinct Tle families. Each family was aligned using the MUSCLE algorithm and phylogenetic trees were generated using the PhyML 3.0 method with bootstrap analysis of 1,000 replicates^{34,35}. Proteins encoded by the genes shown in Fig. 1a were analysed for subcellular localization using the SignalP 3.0 and TMHMM 2.0 servers, and VgrG proteins were identified using blastp^{36,37}. Regions depicted in Fig. 1a were extracted based on boundaries defined by the presence of a *tle*, *tli* or *vgrG* gene. Figure 1b catalytic residues were determined by both PHYRE 2 structural alignment with known lipase enzymes and conservation of those residues within the Tle family alignments. Sequence logos were generated from a manual alignment of conserved catalytic motifs using Geneious software.

Western blot analyses. Whole-cell fractions were prepared as described previously²⁹. Anti-RNA polymerase, anti-VSV-G, anti-β-lactamase, anti-His₆, and anti-cAMP receptor protein (CRP) western blot analyses were performed using previously defined methods^{9,23,38}. To analyse the expression of epitope-tagged Tle2^{VC} and Tle2^{VC}(Ser371Ala) in *V. cholerae*, cells were grown in LB medium at 37 °C to an attenuation (*D*) at 600 nm (*D*_{600 nm}) of 0.5, induced with 0.0002% (w/v) arabinose, and then collected at a final *D*_{600 nm} of 2.0. To analyse the expression of epitope-tagged Tle5^{PA}, Tle5^{PA}(H167R) and Tle5^{PA}(H855R), in *P. aeruginosa*, cells were grown in LB medium supplemented with 1 mM IPTG at 37 °C and collected at a *D*_{600 nm} of 1.0. Subcellular localization of epitope-tagged Tle5^{PA} and Tle5^{PA} in *P. aeruginosa* was performed identically to previous localization studies of Tsi1 and Tsi3 (refs 9, 39). For immunoprecipitation experiments, BL21(DE3) plyS cells co-expressing periplasmic His₆-tagged Tle5^{PA} from a pET22b+ vector and VSV-G-tagged immunity proteins from pScrhaB2 vectors were pelleted and resuspended in lysis buffer (20 mM Tris-Cl, pH 7.5, 50 mM KCl, 8.0% (v/v) glycerol, 1.0% (v/v) Triton, supplemented with DNase I (Roche), lysozyme (Roche) and 200 µM phenylmethylsulphonyl fluoride (PMSF)). Cells were disrupted by sonication and the solution clarified by centrifugation. A sample of supernatant was then taken for analysis of total protein. The remainder of the supernatant was incubated with anti-VSV-G agarose beads (Sigma) for 1 h at 4 °C. Beads were washed four times with immunoprecipitation wash buffer (100 mM NaCl, 25 mM KCl, 0.1% (v/v) Triton, 20 mM Tris-Cl, pH 7.5, and 2% (v/v) glycerol). Proteins were removed from beads with SDS-loading buffer (125 mM Tris, pH 6.8, 2% (w/v) 2-mercaptoethanol, 20% (v/v) glycerol, 0.001% (w/v) bromophenol blue and 4% (w/v) SDS), and analysed by western blot.

Bacterial competition experiments. *Burkholderia* competition experiments were performed as described previously¹⁰. Recipient strains (Fig. 2b, left) or donor strains (Fig. 2b, right) were labelled with a GFP-expression construct integrated into the attTn7 site, allowing the disambiguation of donor and recipient colonies through fluorescence imaging⁴⁰. For *V. cholerae* competition experiments with *E. coli*, both strains were grown to a *D*_{600 nm} of 0.5 in LB medium before being mixed 1:1 by volume. This mixture was then spotted on a nitrocellulose membrane on a 1.5% (w/v) agar LB plate containing 300 mM NaCl and 0.002% (w/v) arabinose. Competitions were incubated for 5 h at 37 °C. Cells were then collected and competitions analysed. Initial and final colony-forming units of *V. cholerae* and *E. coli* were enumerated on LB plates supplemented with rifampin and streptomycin, respectively. For *P. aeruginosa* competitions with *P. putida*, strains were grown overnight on solid LB medium at 37 °C (*P. aeruginosa*) or 30 °C (*P. putida*) and resuspended in water to a *D*_{600 nm} of 0.3. Cells were mixed 1:1 and spotted on 1.5% (w/v) agar SCFM media plates, or inoculated into liquid media of the same. After 23 h of incubation at 23 °C, a temperature previously demonstrated conducive to H2-T6SS and Tle5^{PA} expression under *in vitro* conditions⁴¹, cells were collected and relative numbers of bacteria determined. Both initial and final counts of *P. aeruginosa* and *P. putida* were determined by plate counts. *P. aeruginosa* self-intoxication assays were performed under identical conditions to solid media competition assays, save for the addition of 1 mM IPTG. After 23 h of growth, cells were stained with 5 µg ml⁻¹ propidium iodide in PBS, pH 7.0, for 10 min and washed before fluorescence measurements at an excitation/emission of 535/617 nm. Values shown were corrected for cellular density as measured by *D*_{600 nm}. Competition results for *B. thailandensis* and *P. aeruginosa* experiments are the change in ratio of donor cells to recipient cells, competition results from *V. cholerae* represent the final ratio alone. Data from all competitions were analysed by a two-tailed Student's *t*-test, and data from monoculture experiments

were analysed by a one-tailed Student's *t*-test for a significant increase in propidium iodide staining.

Enzymatic assays of lipase activity. The hydrolysis of polysorbate 20 was measured as described previously⁴². These experiments were performed at 28 °C at a final enzyme concentration of 60 nM in a buffer consisting of 20 mM Tris-Cl, pH 7.2, 100 mM NaCl, 3 mM CaCl₂ and 2% (v/v) polysorbate 20. Fluorescence assays for phospholipase A activity were performed using PED-A1 (*sn*-1-labelled) and PED6 (*sn*-2-labelled) fluorescent substrates according to manufacturer's directions (Invitrogen). Activity of Tle1^{BT} and Tle2^{VC} on these substrates was measured at an enzyme concentration of 300 nM (Tle1^{BT}) or 30 nM (Tle2^{VC}) at 28 °C. For Tle1^{BT}-inhibition assays, immunoprecipitate was obtained as detailed under western blot analyses from *E. coli* DH5 α bearing a pScrhaB2::tli1^{BT}-V expression construct or the equivalent empty vector control, with the modification that proteins were eluted from anti-VSV-G agarose beads by the addition of VSV-G peptide at a concentration of 100 μ g ml⁻¹ and no PMSF was used. After the addition of immunoprecipitate to Tle1^{BT} enzymatic reactions, samples were incubated for 4 min, after which the first reading was normalized to the measurement immediately before treatment. Fluorescent assays for phospholipase D activity were performed by measuring the production of peroxide by choline oxidase through the generation of the fluorescent molecule resorufin from Amplex red reagent (Invitrogen) according to the manufacturer's directions with the following modifications: reactions were performed in a buffer consisting of 50 mM Tris-Cl, pH 7.2, 100 mM NaCl, 5 mM CaCl₂ and 2 mM MgCl₂, and vesicles consisting of equal amounts dioleoylphosphatidylcholine and dioleoylphosphatidylglycerol were used as a substrate at a final reaction concentration of 16.7 μ M for each lipid species. Activity was measured at an enzyme concentration of 130 nM at 28 °C. In all assays fluorescent values were corrected for fluorescence as measured in a buffer-only control.

Competitive lysis assays. The lysis of *P. aeruginosa* reporter strains was determined by the relative partitioning of LacZ to the supernatant. Lysis reporter strains were generated by the chromosomal integration of a previously described miniCTX vector containing *lacZ* under the expression of a constitutive promoter⁴³. Lysis reporter strains and unmarked donor strains were grown overnight on solid LB medium at 37 °C and then resuspended in water to a $D_{600\text{ nm}}$ of 0.3. Donor and recipient strains were mixed 1:1 and spotted on 1.5% (w/v) agar SCFM plates supplemented with 1 mM IPTG and incubated at 23 °C for 23 h. Relative levels of supernatant LacZ activity as compared to total LacZ activity were then determined as previously described²⁰. Data were analysed using a two-tailed Student's *t*-test.

Microscopic analyses of interbacterial competitions. Time-lapse fluorescence microscopy sequences were acquired with a Nikon Ti-E inverted microscope fitted with a $\times 60$ oil objective, automated focusing (Perfect Focus System, Nikon), a Xenon light source (Sutter Instruments), a CCD camera (Clara series, Andor), and a custom environmental chamber. NIS Elements (Nikon) was used for automated image acquisition. Overnight cultures of recipient (*B. thailandensis* Δ BTH_I2698-I2703 attTn7::gfp) and donor (either *B. thailandensis* wild-type, Δ BTH_I2698 Δ BTH_I2701-3, or Δ BTH_I2598) strains were mixed 1:1 and diluted twofold with LB medium. The resulting bacterial suspension (~ 2 μ l) was spotted onto growth pads made with LB medium, 2.5% (w/v) agarose, 0.2% (w/v) sodium nitrate, and 2.5 μ g ml⁻¹ propidium iodide. Automated image acquisition was performed at 5-min intervals for 6–8 h at 30 °C. Cell identification, cell linking, and donor-control analyses were performed using customized Matlab-based software (2012a, Mathworks) as described previously¹³. Donor (unlabelled) and recipient (GFP-labelled) populations were identified using an empirically determined green fluorescence gate. A propidium iodide uptake event was defined as the first frame in which a cell achieved an empirically determined mean red fluorescence intensity threshold. Counting error was calculated as the square root of measurable events. Results represent two fields of view from a single experiment; each experiment was independently repeated at least three times. Videos generated from cropped regions of the three growth competition experiments depicted are provided (Supplementary Videos 1–3).

Protein purification. For purification, Tle proteins were expressed from pET28b::His₆-MBP-TEV-His₆ in Shuffle T7 pLysY Express cells (New England Biolabs). Proteins were purified to homogeneity using nickel chromatography followed by size-exclusion chromatography using previously reported methods, with the exception that reducing agents were excluded⁴⁴.

Lipidomic analyses. Wild-type and *tli5* mutant *P. aeruginosa* strains were grown as 20 individual 10- μ l spots on 1.5% (w/v) agar SCFM plates for 23 h at 23 °C. These spots were then resuspended in PBS and lipids were extracted using the Bligh–Dyer method⁴⁵. Purified lipid samples were analysed for phosphatidylethanolamine, phosphatidylcholine, phosphatidylglycerol and phosphatidic acid content by the Kansas State University Lipidomics Research Center. An automated electrospray ionization-tandem mass spectrometry approach was used, and data

acquisition and analysis were carried out as described previously^{46,47} with modifications. The lipid samples were dissolved in 1 ml chloroform. An aliquot of 50 μ l of extract in chloroform was used. Precise amounts of internal standards, obtained and quantified as previously described⁴⁸, were added in the following quantities (with some small variation in amounts in different batches of internal standards): 0.6 nmol didodecylphosphatidylcholine (di12:0-PC), 0.6 nmol di24:1-phosphatidylcholine (PC), 0.6 nmol 13:0-lysoPC, 0.6 nmol 19:0-lysoPC, 0.3 nmol di12:0-phosphatidylethanolamine (PE), 0.3 nmol di23:0-PE, 0.3 nmol 14:0-lysoPE, 0.3 nmol 18:0-lysoPE, 0.3 nmol di14:0-phosphatidylglycerol (PG), 0.3 nmol di20:0(phytanoyl)-PG, 0.3 nmol di14:0-phosphatidic acid (PA), and 0.3 nmol di20:0(phytanoyl)-PA. The sample and internal standard mixture was combined with solvents, such that the ratio of chloroform:methanol:300-mM ammonium acetate in water was 300:665:35, and the final volume was 1.4 ml. Unfractionated lipid extracts were introduced by continuous infusion into the ESI source on a triple quadrupole tandem mass spectrometer (MS/MS; 4000QTrap, Applied Biosystems). Samples were introduced using an autosampler (LC Mini PAL, CTC Analytics AG) fitted with the required injection loop for the acquisition time and presented to the electrospray ionization (ESI) needle at 30 μ l min⁻¹. Sequential precursor and neutral loss scans of the extracts produce a series of spectra with each spectrum revealing a set of lipid species containing a common head group fragment. Lipid species were detected with the following scans: phosphatidylcholine and lysoPC, [M + H]⁺ ions in positive ion mode with precursor of 184.1 (Pre 184.1); phosphatidylethanolamine and lysoPE, [M + H]⁺ ions in positive ion mode with neutral loss of 141.0 (NL 141.0); phosphatidylglycerol, [M + NH₄]⁺ in positive ion mode with NL 189.0 for phosphatidylglycerol; and phosphatidic acid, [M + NH₄]⁺ in positive ion mode with NL 115.0. The collision gas pressure was set at 2 (arbitrary units). The collision energies, with nitrogen in the collision cell, were +28 V for phosphatidylethanolamine, +40 V for phosphatidylcholine, +25 V for phosphatidic acid, and +20 V for phosphatidylglycerol. Declustering potentials were +100 V for all lipids. Entrance potentials were +15 V for phosphatidylethanolamine and +14 V for phosphatidylcholine, phosphatidic acid and phosphatidylglycerol. Exit potentials were +11 V for phosphatidylethanolamine and +14 V for phosphatidylcholine, phosphatidic acid and phosphatidylglycerol. The scan speed was 50 or 100 μ s⁻¹. The mass analysers were adjusted to a resolution of 0.7 U full-width at half height. For each spectrum, 9–150 continuum scans were averaged in multiple channel analyser mode. The source temperature (heated nebulizer) was 100 °C, the interface heater was on, +5.5 kV were applied to the electrospray capillary, the curtain gas was set at 20 (arbitrary units), and the two ion source gases were set at 45 (arbitrary units). The background of each spectrum was subtracted, the data were smoothed, and peak areas integrated using a custom script and Applied Biosystems Analyst software, and the data were isotopically deconvoluted. The first set of mass spectra was acquired on the internal standard mixture only. Peaks corresponding to the target lipids in these spectra were identified and molar amounts calculated in comparison to the two internal standards on the same lipid class. To correct for chemical or instrumental noise in the samples, the molar amount of each lipid metabolite detected in the 'internal standards only' spectra was subtracted from the molar amount of each metabolite calculated in each set of sample spectra. The data from each 'internal standards only' set of spectra was used to correct the data. Values expressed are the percentage of the total polar lipid signal detected. Statistical significance analysed by a two-tailed Student's *t*-test.

- Chandler, J. R. *et al.* Mutational analysis of *Burkholderia thailandensis* quorum sensing and self-aggregation. *J. Bacteriol.* **191**, 5901–5909 (2009).
- Vaitkevicius, K. *et al.* A *Vibrio cholerae* protease needed for killing of *Caenorhabditis elegans* has a role in protection from natural predator grazing. *Proc. Natl Acad. Sci. USA* **103**, 9280–9285 (2006).
- Valeru, S. P. *et al.* Role of melanin pigment in expression of *Vibrio cholerae* virulence factors. *Infect. Immun.* **77**, 935–942 (2009).
- Rietsch, A., Vallet-Gely, I., Dove, S. L. & Mekalanos, J. J. ExsE, a secreted regulator of type III secretion genes in *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA* **102**, 8006–8011 (2005).
- Nelson, K. E. *et al.* Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**, 799–808 (2002).
- Horton, R. M. *et al.* Gene splicing by overlap extension. *Methods Enzymol.* **217**, 270–279 (1993).
- Korotkov, K. V. & Hol, W. G. Crystal structure of the pilin from the enterohemorrhagic *Escherichia coli* type II secretion system. *J. Struct. Biol.* <http://dx.doi.org/10.1016/j.jsb.2013.02.013> (28 February 2013).
- Cardona, S. T. & Valvano, M. A. An expression vector containing a rhamnose-inducible promoter provides tightly regulated gene expression in *Burkholderia cenocepacia*. *Plasmid* **54**, 219–228 (2005).
- Hsu, F., Schwarz, S. & Mougous, J. D. TagR promotes PpkA-catalysed type VI secretion activation in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **72**, 1111–1125 (2009).

30. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
31. Schell, M. A. *et al.* Type VI secretion is a major virulence determinant in *Burkholderia mallei*. *Mol. Microbiol.* **64**, 1466–1485 (2007).
32. Donnenberg, M. S. & Kaper, J. B. Construction of an *eae* deletion mutant of enteropathogenic *Escherichia coli* by using a positive-selection suicide vector. *Infect. Immun.* **59**, 4310–4317 (1991).
33. Kelley, L. A. & Sternberg, M. J. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**, 363–371 (2009).
34. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
35. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
36. Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795 (2004).
37. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2**, 953–971 (2007).
38. Balsalobre, C. *et al.* Release of the type I secreted alpha-haemolysin via outer membrane vesicles from *Escherichia coli*. *Mol. Microbiol.* **59**, 99–112 (2006).
39. Imperi, F. *et al.* Analysis of the periplasmic proteome of *Pseudomonas aeruginosa*, a metabolically versatile opportunistic pathogen. *Proteomics* **9**, 1901–1915 (2009).
40. Schwarz, S. *et al.* *Burkholderia* type VI secretion systems have distinct roles in eukaryotic and bacterial cell interactions. *PLoS Pathog.* **6**, e1001068 (2010).
41. Termine, E. & Michel, G. P. Transcriptome and secretome analyses of the adaptive response of *Pseudomonas aeruginosa* to suboptimal growth temperature. *Int. Microbiol.* **12**, 7–12 (2009).
42. von Tigerstrom, R. G. & Stelmaschuk, S. The use of Tween 20 in a sensitive turbidimetric assay of lipolytic enzymes. *Can. J. Microbiol.* **35**, 511–514 (1989).
43. Vance, R. E., Rietsch, A. & Mekalanos, J. J. Role of the type III secreted exoenzymes S, T, and Y in systemic spread of *Pseudomonas aeruginosa* PAO1 *in vivo*. *Infect. Immun.* **73**, 1706–1713 (2005).
44. Mougous, J. D. *et al.* Identification, function and structure of the mycobacterial sulfotransferase that initiates sulfolipid-1 biosynthesis. *Nature Struct. Mol. Biol.* **11**, 721–729 (2004).
45. Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, 911–917 (1959).
46. Brügger, B., Erben, G., Sandhoff, R., Wieland, F. T. & Lehmann, W. D. Quantitative analysis of biological membrane lipids at the low picomole level by nano-electrospray ionization tandem mass spectrometry. *Proc. Natl Acad. Sci. USA* **94**, 2339–2344 (1997).
47. Devaiah, S. P. *et al.* Quantitative profiling of polar glycerolipid species from organs of wild-type Arabidopsis and a *PHOSPHOLIPASE Dα1* knockout mutant. *Phytochemistry* **67**, 1907–1924 (2006).
48. Weltri, R. *et al.* Profiling membrane lipids in plant stress response. *J. Biol. Chem.* **277**, 31994–32002 (2002).

Induction of pathogenic T_H17 cells by inducible salt-sensing kinase SGK1

Chuan Wu^{1*}, Nir Yosef^{1,2*}, Theresa Thalhamer^{1†*}, Chen Zhu¹, Sheng Xiao¹, Yasuhiro Kishi¹, Aviv Regev^{2,3} & Vijay K. Kuchroo^{1,2}

T_H17 cells (interleukin-17 (IL-17)-producing helper T cells) are highly proinflammatory cells that are critical for clearing extracellular pathogens and for inducing multiple autoimmune diseases¹. IL-23 has a critical role in stabilizing and reinforcing the T_H17 phenotype by increasing expression of IL-23 receptor (IL-23R) and endowing T_H17 cells with pathogenic effector functions^{2,3}. However, the precise molecular mechanism by which IL-23 sustains the T_H17 response and induces pathogenic effector functions has not been elucidated. Here we used transcriptional profiling of developing T_H17 cells to construct a model of their signalling network and nominate major nodes that regulate T_H17 development. We identified serum glucocorticoid kinase 1 (SGK1), a serine/threonine kinase⁴, as an essential node downstream of IL-23 signalling. SGK1 is critical for regulating IL-23R expression and stabilizing the T_H17 cell phenotype by deactivation of mouse Foxo1, a direct repressor of IL-23R expression. SGK1 has been shown to govern Na⁺ transport and salt (NaCl) homeostasis in other cells^{5–8}. We show here that a modest increase in salt concentration induces SGK1 expression, promotes IL-23R expression and enhances T_H17 cell differentiation *in vitro* and *in vivo*, accelerating the development of autoimmunity. Loss of SGK1 abrogated Na⁺-mediated T_H17 differentiation in an IL-23-dependent manner. These data demonstrate that SGK1 has a critical role in the induction of pathogenic T_H17 cells and provide a molecular insight into a mechanism by which an environmental factor such as a high salt diet triggers T_H17 development and promotes tissue inflammation.

To determine the molecular mechanisms by which naive T cells develop into effector T_H17 cells, we measured genome-wide messenger RNA expression profiles using microarrays along 18 time points over 72 h, following the *in vitro* exposure of naive T cells to T_H17 polarizing conditions (transforming growth factor β1 (TGF-β1) with IL-6). To examine the role of IL-23 in T_H17 development, we added IL-23 at the late time points (48–72 h) and monitored the transcriptional response in both wild-type and *Il23r*^{−/−} cells. We ranked the genes according to their extent of induction in cells treated with TGF-β1 and IL-6 (relative to non-polarized activated T cells) and repression in *Il23r*^{−/−} cells (relative to wild-type cells) (Methods, Fig. 1a and Supplementary Table 1). Murine *Sgk1* was one of the top ranking genes, whose transcriptional regulation was strongly associated with both IL-23R signalling and T_H17 cell differentiation (Fig. 1a). Quantitative polymerase chain reaction (qPCR) analysis showed that *Sgk1* is induced at low levels by TGF-β1 (induced regulatory T (iT_{reg}) cells), and not induced in other T cell subsets (T_H0, T_H1, T_H2). As expected, it is most highly expressed under T_H17 differentiation conditions (Fig. 1b). *Sgk1* expression is strongly induced during the first 2 h after stimulation of naive T cells under T_H17-polarizing conditions. This is followed by a sharp decline by 10 h to a steady expression level that is still substantially higher than in the control population (Fig. 1c and Supplementary Fig. 1a). Furthermore, *Sgk1* expression is specifically induced and

maintained by exposure to IL-23 (Fig. 1c and Supplementary Fig. 1b). Although *Il23r*^{−/−} T cells initially produce *Sgk1* mRNA, they cannot sustain this expression (Supplementary Fig. 1b, c). Finally, the kinase activity of SGK1 is also significantly higher in T_H17 cells than in other T cell subsets (Supplementary Fig. 1d), and restimulation of differentiated T_H17 cells with IL-23 elevates further SGK1 kinase activity (Supplementary Fig. 1e). Thus, IL-23 signalling is critical for maintaining *Sgk1* expression during T_H17 cell differentiation.

Network analysis of the transcriptional changes in *Il23r*^{−/−} T cells using the ANAT software⁹ singled out SGK1 as a potential nodal point downstream of IL-23R signalling. Based on a curated database of protein–protein interactions (PPIs), we constructed a network model that connects known proteins of the IL-23R signalling pathway (Methods) to the transcription factors whose function is dysregulated in *Il23r*^{−/−} cells (Methods, Fig. 1d and Supplementary Fig. 1f). We ranked the network's nodes based on a centrality measure, defined as the fraction of IL-23R-affected transcription factors downstream of that node in the network (Methods and Supplementary Table 1). SGK1 was the highest-ranking node (Supplementary Fig. 1g), suggesting that it acts both as a transcriptional target of IL-23R signalling and as a kinase that may mediate the transcriptional effects of the pathway.

Using *Sgk1*^{−/−} mice, we studied the impact of loss of SGK1 on T_H17 differentiation *in vitro*. We observed no abnormality of SGK1-deficient T cells during primary differentiation into T_H17 cells (Fig. 1e). However, *Sgk1*^{−/−} T_H17 cells restimulated with IL-23 showed impaired IL-17 production (Fig. 1e and Supplementary Fig. 2b). Memory *Sgk1*^{−/−} T cells also showed a defect in IL-17 production upon IL-23 stimulation, but not under stimulation with TGF-β1 and IL-6 (Supplementary Fig. 2a). To study the function of SGK1 specifically in IL-17-producing T cells that carry the CD4 antigen (CD4⁺ T cells), we generated *Il17f*^{Cre}*Sgk1*^{fl/fl} mice in which SGK1 was deleted in cells producing IL-17E, enabling us to analyse the function of SGK1 in the maintenance of T_H17 phenotype. *Il17f*^{Cre}*Sgk1*^{fl/fl} T cells also showed no defect in primary T_H17 differentiation, but displayed reduced IL-17 production when restimulated with IL-23 (Fig. 1f and Supplementary Fig. 2c). IL-23R expression was also significantly reduced in *Sgk1*^{−/−} T cells (Fig. 1g). Thus, loss of SGK1 does not affect primary T_H17 differentiation, but profoundly affects their stability and IL-23R expression. One possible explanation for the dispensability of SGK1 during primary T_H17 differentiation is redundancy with other kinases, such as its homologue AKT¹⁰. However, SGK1 seems to be indispensable for IL-23R-dependent stability and maintenance of T_H17 cells.

Microarray analysis of *Sgk1*^{−/−} versus wild-type T_H17 cells restimulated with IL-23 showed a significant overlap in differentially expressed genes with the *Il23r*^{−/−} versus wild-type IL-23-restimulated T_H17 cell profiles, supporting further the functional relatedness of the SGK1 and IL-23R pathways (Fisher exact test, *P* < 0.001) (Fig. 1h and Supplementary Fig. 2d). Consistent with this, genes downregulated in *Sgk1*^{−/−} cells are significantly enriched (Fisher exact test, *P* < 0.001)

¹Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ³Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02140, USA. †Present address: Department of Molecular Biology, University of Salzburg, Hellbrunnerstrasse 34, 5020 Salzburg, Austria.

*These authors contributed equally to this study.

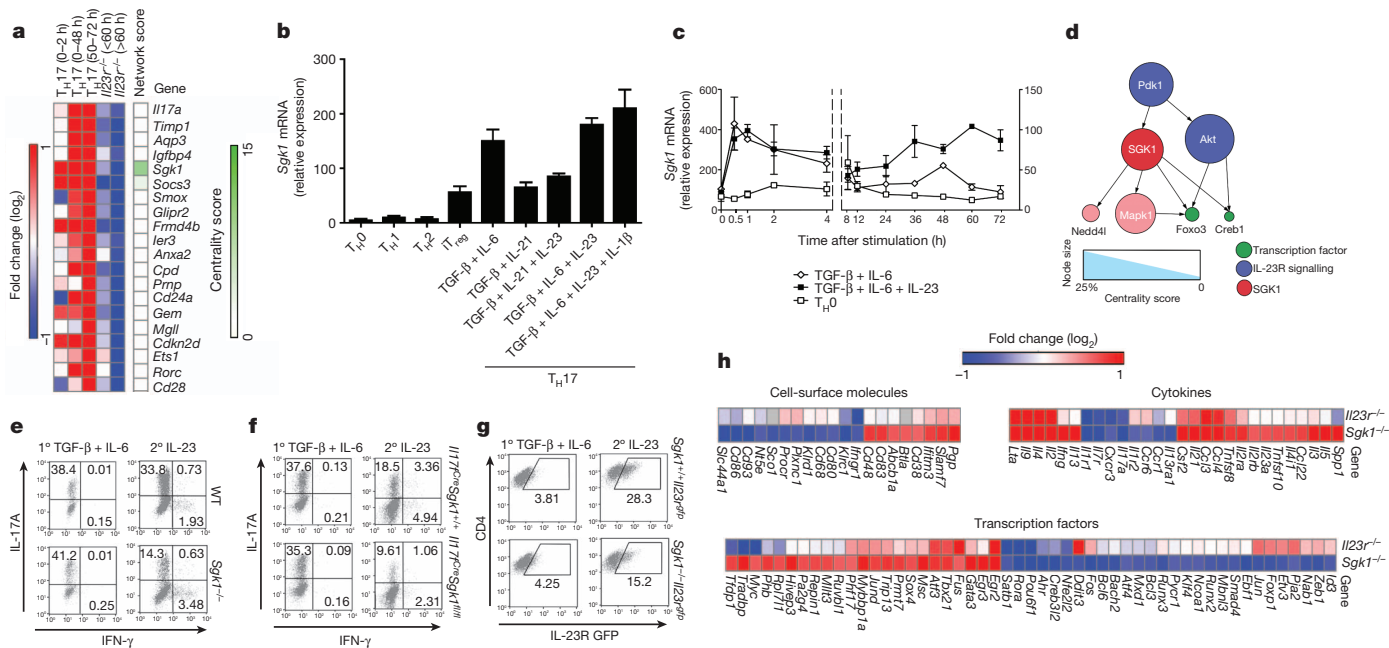


Figure 1 | SGK1 is specifically induced in T_H17 cells and is important for their maintenance. **a**, Top candidate genes ranked by their average of fold-increase in T_H17 conditions (TGF- β 1 with IL-6 compared to T_H0) and fold decrease in $Il23r^{-/-}$ (knockout versus wild-type cells, TGF- β 1, IL-6 and IL-23 condition). The centrality score of a given protein is the percentage of IL-23R-affected transcription factors downstream of that protein in the network (Methods). **b**, *Sgk1* mRNA expression in different T cell subsets. **c**, Kinetic analysis of *Sgk1* gene expression in activated naive wild-type $CD4^+$ T cells differentiated with TGF- β , IL-6 and IL-23. **d**, IL-23R PPI network model (this is an enlargement of the SGK1 sub-network from the full network of Supplementary Fig. 1f). Nodes

for genes that are upregulated in wild-type T_H17 cells compared to other T cell subsets¹¹ (Methods and Supplementary Fig. 2e). Selected genes were confirmed by qPCR analysis (Supplementary Fig. 2f). Genes from several other pathways are also enriched (over- or underexpressed) (Supplementary Table 2), including cell cycle and proliferation, which may be related to the known role of SGK1 as a regulator of proliferation and apoptosis^{7,8,10}. Although our analysis strongly associates SGK1 with the T_H17 program, genes important for development and function of other T cell subsets, such as *Ifng*, *Tbx21* or *Gata3* were also dysregulated in $Sgk1^{-/-}$ cells, suggesting possible additional effects of this kinase in other T cell subsets.

To determine the role of SGK1 *in vivo*, we immunized $Cd4^{Cre}Sgk1^{fl/fl}$ mice with myelin oligodendrocyte glycoprotein (MOG) peptide 35–55 (MOG_{35–55}) to induce experimental autoimmune encephalomyelitis (EAE). SGK1-deficient mice exhibited significantly reduced EAE incidence and severity. IL-17 production from infiltrated $CD4^+$ T cells in different organs of SGK1-deficient mice was also reduced, whereas interferon- γ (IFN- γ) levels were unaffected (Fig. 2a and Supplementary Fig. 3a). When we restimulated the isolated T cells from immunized mice with IL-23 in the presence of MOG_{35–55}, the SGK1-deficient T cells also showed impaired IL-17 but normal IFN- γ production (Supplementary Fig. 3b, c). Next, using *Il23r^{gfp}* reporter mice, we observed reduced IL-23R-GFP (green fluorescent protein) expression on infiltrating $CD4^+$ T cells in different organs of SGK1-deficient mice undergoing EAE (Supplementary Fig. 4a). Similar to the response of $Cd4^{Cre}Sgk1^{fl/fl}$ mice, reduced T_H17 differentiation and disease severity were also observed in *Il17f^{Cre}Sgk1^{fl/fl}* mice during EAE (Fig. 2b). In addition, to exclude any effects of SGK1-deficient bystander cells, we transferred purified *Il17f^{Cre}Sgk1^{fl/fl}* $CD4^+$ T cells into *Rag2^{-/-}* mice and induced EAE. Mice that received SGK1-deficient T cells developed less severe disease compared to mice that received wild-type T cells (Supplementary Fig. 4b).

are sized proportionally to their centrality score. **e–g**, Naive $CD4^+$ T cells from $Sgk1^{-/-}$ (**e**), *Il17f^{Cre}Sgk1^{fl/fl}* (**f**) or *Sgk1^{-/-}Il23r^{gfp}* (**g**) and control mice were differentiated to T_H17 cells with TGF- β 1 and IL-6 (left) or restimulated with IL-23 (right). IL-17 and IFN- γ or IL-23R (GFP) expression were assessed. Numbers in the graphs indicate the percentage of cells in that quadrant. **h**, Heat map displaying microarray data, fold change of selected gene subsets in the two experimental settings: $Sgk1^{-/-}$ versus wild-type, and *Il23r^{-/-}* versus wild-type T_H17 cells (TGF- β 1 and IL-6, restimulated with IL-23). Only genes with a significant fold change in the $Sgk1^{-/-}$ T_H17 cells are presented. Data are representative of at least two independent experiments. Error bars, s.d.

To determine the reason for fewer T_H17 cells being found in SGK1-deficient mice, we transferred purified GFP⁺ cells from differentiated *Cd4^{Cre}Sgk1^{fl/fl}Il17a^{gfp}* or control T_H17 cells to congenic Ly5.1 mice and traced the IL-17 GFP⁺ cells in different organs after immunization with MOG_{35–55} (Fig. 2c). Starting with the same number of $CD4^+$ IL-17⁺ T cells, we found that 7 and 12 days after transfer, SGK1-deficient T_H17 cells failed to maintain IL-17 production, particularly in the central nervous system (CNS) (Fig. 2d and Supplementary Fig. 4c). Next, we crossed *Il17f^{Cre}Rosa26R^{eYFP}* mice onto the SGK1-deficient background, and analysed the expression of IL-17 in T cells that had turned on the *Il17f* gene as determined by enhanced yellow fluorescent protein (eYFP) expression. We induced EAE in these mice and analysed the frequency of eYFP⁺ cells producing IL-17 in infiltrating $CD4^+$ T cells in the lymph nodes and CNS. The $Sgk1^{-/-}$ reporter mice showed a smaller proportion of $CD4^+$ eYFP⁺ T cells in both organs. Furthermore, there was a dramatic loss of IL-17 expression by eYFP⁺ T cells in the SGK1-deficient mice, indicating that T_H17 cells could not stably retain IL-17 production during EAE (Fig. 2e and Supplementary Fig. 4d).

To understand better the molecular role of SGK1 in T_H17 cells, we conducted another network analysis, using PPI data to connect SGK1 to the transcription factors whose activity is dysregulated in $Sgk1^{-/-}$ T_H17 cells (Methods). The analysis suggested Foxo1 as one of the highest-ranking nodes downstream of SGK1 (Fig. 3a, Supplementary Table 1 and Supplementary Fig. 5a). Foxo1 phosphorylation by SGK1 in adipocytes has been shown previously to lead to its deactivation and translocation from the nucleus to the cytoplasm¹². Consistent with this observation, we found that SGK1 phosphorylates Foxo1 (Supplementary Fig. 5b). Immunoblot analysis of $Sgk1^{-/-}$ T_H17 cells restimulated with IL-23 confirmed that there is not only reduced phosphorylation of Foxo1 in the nucleus but increased mRNA and protein expression of Foxo1 (Fig. 3b, c), suggesting that compromised phosphorylation of Foxo1 can result in its own transcriptional upregulation. It has been

shown previously that Foxo1 can regulate its own expression¹³ and we have found that Foxo1 binds to a site located about 1 kilobase (kb) upstream of the first exon in the *Foxo1* locus (Supplementary Fig. 6a).

Transfection of a *Foxo1* luciferase reporter in the presence of Foxo1 led to increased luciferase activity (Supplementary Fig. 6b), whereas increasing expression of SGK1 in the presence of Foxo1 resulted in a dose-dependent decrease in reporter activity, suggesting that SGK1 inhibited Foxo1-mediated transactivation of its own promoter (Fig. 3d).

To decipher the consequences of Foxo1 expression on T_H17 cell development, we used *Foxo1*^{-/-} CD4⁺ memory T cells and observed higher expression of *Il23r* and *Il17a* in these cells than in wild-type cells, indicating that Foxo1 may act as a repressor of T_H17 cell development and of IL-23 signalling (Fig. 3e and Supplementary Fig. 6c). We also found potential binding sites of Foxo1 located about 1 kb upstream of the first exon of the *Il23r* locus by chromatin immunoprecipitation (ChIP)-PCR (Supplementary Fig. 6d). Moreover, there is significantly enriched binding of Foxo1 on the *Il23r* promoter region in *Sgk1*^{-/-} cells compared to that in wild-type T cells, indicating enhanced suppression of *Il23r* transcription in the absence of SGK1 (Fig. 3f). Retinoic-acid-receptor-related orphan receptor γ t (ROR γ t) has been suggested to be the master transcription factor of T_H17 development and ChIP-seq¹⁴ (ChIP coupled with high-throughput DNA sequencing) and our ChIP-PCR analysis confirmed that IL-23R is one of the targets of ROR γ t (Supplementary Fig. 6e). Indeed, we observed that the *Il23r* promoter is transactivated by ROR γ t in IL-23-restimulated T_H17 cells and it can be inhibited by Foxo1 in a dose-dependent manner (Supplementary Fig. 6f, g). While Foxo1 inhibited ROR γ t-mediated *Il23r* expression, co-expression of SGK1 together with ROR γ t and Foxo1 abrogated the suppressive effects of Foxo1 and rescued *Il23r*

promoter transcriptional activity (Fig. 3g). Additionally, the inhibition of *Il23r* transcription by a phosphorylation-insensitive triple alanine mutant of Foxo1, Foxo1 AAA, was not reduced in the presence of SGK1 (Supplementary Fig. 6h). Furthermore, we observed an endogenous Foxo1–ROR γ t interaction in primary T_H17 cells (Fig. 3h). These data support a model in which some of the effects of SGK1 are due to phosphorylation of Foxo1, which may be a key step in relieving ROR γ t from Foxo1-mediated inhibition, enhancing the expression of IL-23R.

SGK1 has been reported to act as a mediator for sodium homeostasis. It can be induced by exogenous sodium chloride and is one of the major kinases that regulates Na⁺ intake by phosphorylation of epithelial sodium channels (ENaCs)^{4,5}. Considering the defects in T_H17 development in *Sgk1*^{-/-} mice, this raised the hypothesis that increasing sodium concentration may affect the T_H17 cell phenotype through SGK1. To test this, we first activated naive T cells in the presence of additional NaCl, but in the absence of any polarizing cytokines. Microarray analysis of these NaCl-treated cells showed a significant upregulation of *Sgk1* and of multiple other genes associated with T_H17 development (Fisher exact test; $P < 0.001$; Supplementary Table 2 and Supplementary Fig. 7a), which we confirmed by qPCR analysis of selected genes (Supplementary Fig. 7b). We also found increased mRNA and protein levels of IL-17 and IL-23R with additional NaCl under various T_H17 polarizing conditions (Fig. 4a, b and Supplementary Fig. 7c). Furthermore, a sodium-induced increase in T_H17 development and IL-23R expression was not observed in SGK1-deficient T cells, specifically in the context of IL-23–IL-23R signalling (Fig. 4c, d). Importantly, culturing cells with mannitol did not alter T_H17 cell differentiation, excluding the possibility that the T_H17 program is initiated simply by the alteration of osmotic pressure (Supplementary Fig. 7d).

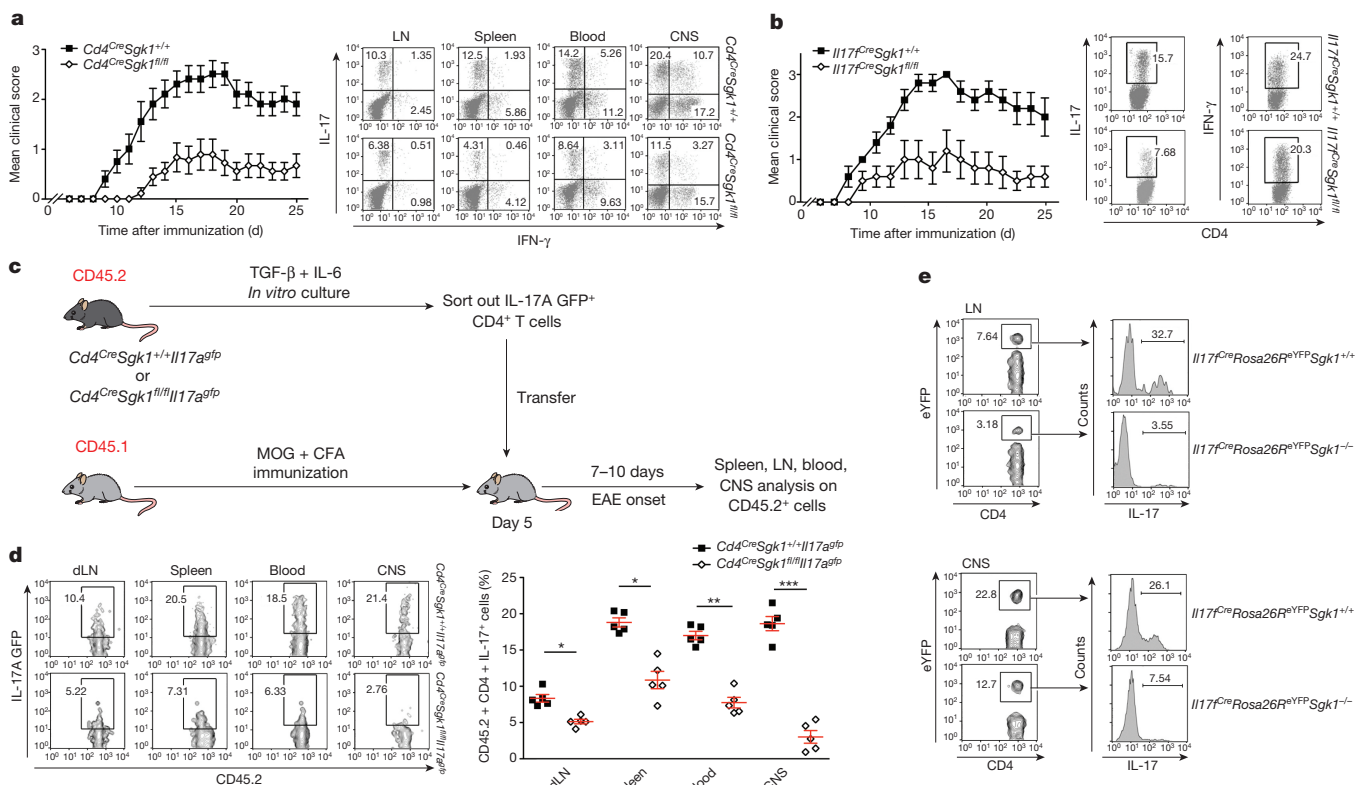


Figure 2 | SGK1-deficient mice are resistant to EAE, owing to a defect in maintaining the T_H17 phenotype. **a**, EAE development in *Cd4^{Cre}Sgk1^{+/+}* and *Cd4^{Cre}Sgk1^{fl/fl}* mice (left), and IL-17 and IFN- γ secretion by CD4⁺ T cells isolated from indicated organs at the peak of disease (right) ($n = 12$). **b**, EAE development in *Il17f^{Cre}Sgk1^{+/+}* and *Il17f^{Cre}Sgk1^{fl/fl}* mice (left), and IL-17 and IFN- γ secretion by CD4⁺ T cells within the CNS (right) ($n = 10$). **c**, Schematic illustration of adoptive transfer experiments shown in **d**. **d**, IL-17 production

from the donor CD4⁺ cells collected from the indicated organs 12 days after transfer; representative histograms (left) and quantification of the flow cytometry data (right; means and s.d. are shown in red, $n = 10$). dLN, draining lymph nodes. **e**, IL-17A production by CD4⁺ eYFP⁺ T cells isolated from lymph nodes or CNS of wild-type and SGK1-deficient *Il17f^{Cre}Rosa26R^{eYFP}Sgk1^{+/+}* and *Il17f^{Cre}Rosa26R^{eYFP}Sgk1^{-/-}* mice 17 days after MOG_{35–55}–CFA immunization ($n = 10$). * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$ (Student's t -test). Error bars, s.d.

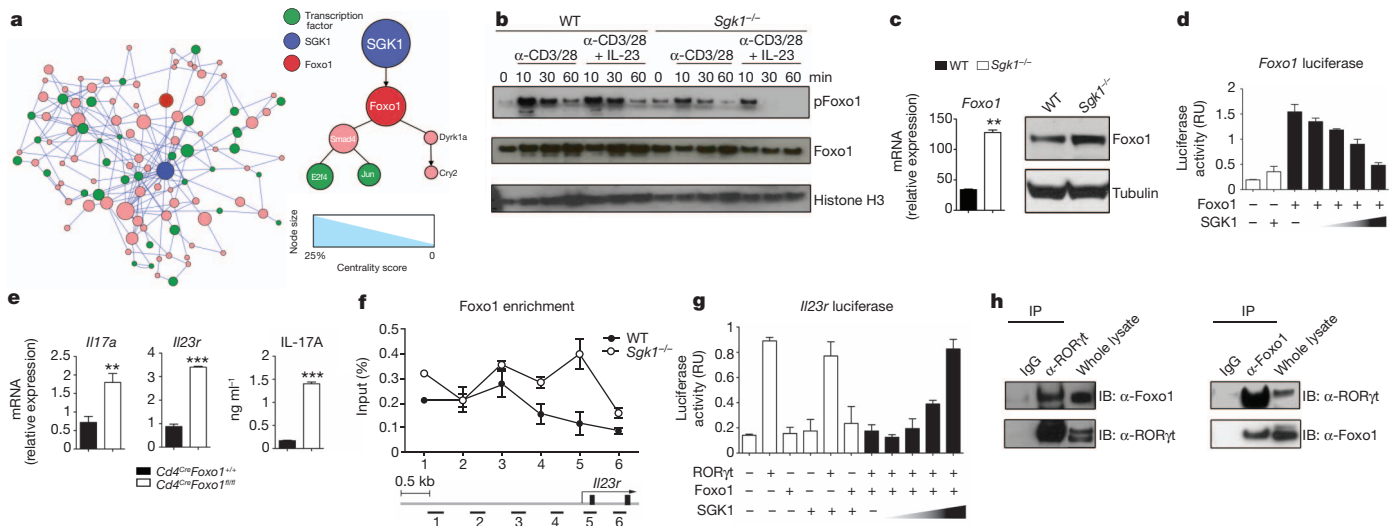


Figure 3 | SGK1 signalling promotes IL-23R expression through phosphorylation of Foxo1. **a**, SGK1 PPI network model. Left, network composed of all the protein nodes with a *P* value of under 0.0001 (Methods); right, enlargement of the Foxo1 sub-network. Nodes are sized relative to their centrality score. In the subnetwork, directed edges (arrows) from one protein to another correspond to post-translational modifications of the second protein by the first. Non-directed edges (lines without arrowheads) correspond to PPIs between one protein and another with no known directionality. **b**, Phosphorylated Foxo1 (p-Foxo1; phosphorylation site Ser 256) and total Foxo1 levels were assessed in nuclear extracts after restimulation of wild-type (WT) and *Sgk1*^{-/-} T_H17 cells. **c**, Levels of mRNA (left) and protein (right) of Foxo1 were analysed 3 days after IL-23 restimulation of differentiated T_H17 cells. **d**, HEK293T cells were transfected with a *Foxo1* promoter-driven luciferase reporter along with the indicated plasmids, and promoter activity was assessed. **e**, Memory CD4⁺ T cells

from *Foxo1*^{-/-} mice were stimulated for 24 h with PMA (phorbol myristate acetate) and ionomycin, and levels of IL-17A and IL-23R expression were determined by qPCR or ELISA (enzyme-linked immunosorbent assay) (IL-17A). **f**, The binding of Foxo1 to the *Il23r* promoter in wild-type and *Sgk1*^{-/-} IL-23-restimulated T_H17 cells was assayed by ChIP-PCR. Six horizontal bars represent the locations of the Foxo1-binding sites on the *Il23r* locus detected by real-time PCR. Two thick vertical bars represent exons 1 and 2 of the *Il23r* locus. **g**, *Il23r* promoter activity was measured in HEK293T cells transfected with an *Il23r* promoter-driven luciferase reporter along with the indicated plasmids. **h**, Immunoprecipitation (IP) (control IgG (immunoglobulin-γ), anti-RORγt or anti-Foxo1) of lysates of wild-type IL-23-restimulated T_H17 cells, followed by immunoblot (IB) analysis with indicated antibodies. ***P* < 0.01 and ****P* < 0.001 (Student's *t*-test). Data are representative of three independent experiments. RU, relative units. Error bars, s.d.

Recent studies have shown that different components in the daily diet and gut microbiota can strongly affect the frequency of effector T cells in the gut^{15,16}. Furthermore, previous data indicate that molecules related to sodium homeostasis can influence T_H17 cell responses^{17,18}. To understand further the effect of NaCl on T_H17 cell generation

in vivo, we fed a high salt diet (HSD) to wild-type or *Cd4*^{Cre}*Sgk1*^{fl/fl} mice. After 3 weeks on HSD, we observed that un-immunized wild-type mice showed a marked increase in the frequency of T_H17 cells in the lamina propria, whereas no notable changes were observed in the mesenteric lymph nodes or spleen. Conversely, SGK1-deficient mice

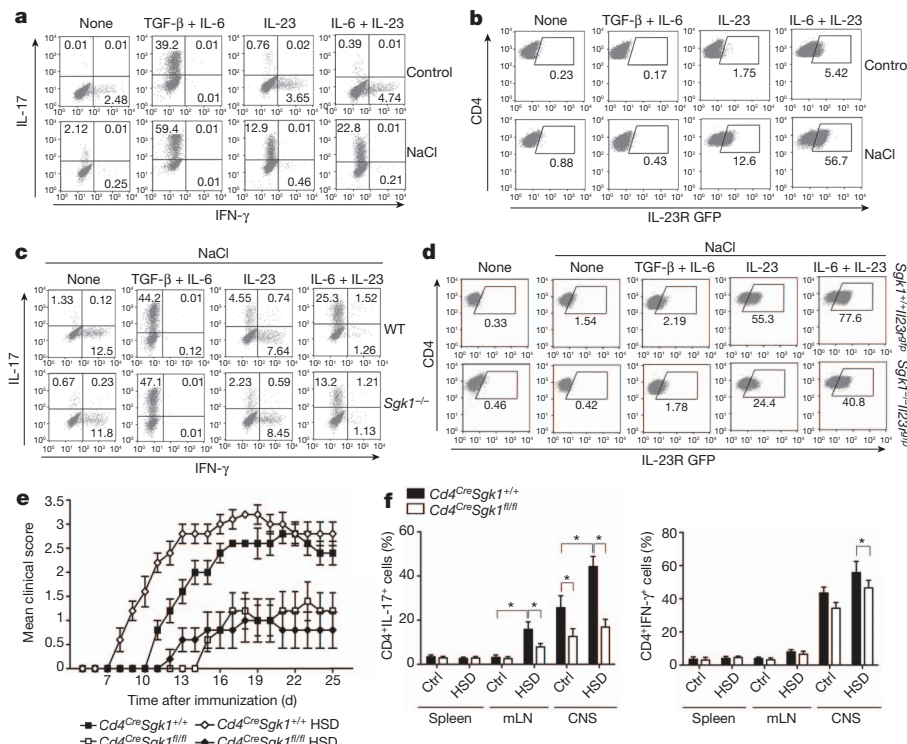


Figure 4 | NaCl potentiates T_H17 cell differentiation *in vitro* and *in vivo*, enhancing EAE induction. **a**, IL-17 and IFN-γ production by naive wild-type CD4⁺ T cells stimulated with indicated cytokines for 3 days with or without 40 mM NaCl. Ctrl, control. **b**, IL-23R (GFP) expression in *Il23r*^{gfp} CD4⁺ T cells stimulated under conditions described in **a**. **c**, IL-17 and IFN-γ production by wild-type (WT) and *Sgk1*^{-/-} CD4⁺ T cells stimulated under conditions described in **a**. **d**, IL-23R (GFP) expression in *Sgk1*^{-/-}*Il23r*^{gfp} and control CD4⁺ T cells stimulated under conditions described in **c**. **e**, Clinical scores of EAE in *Cd4*^{Cre}*Sgk1*^{+/+} and *Cd4*^{Cre}*Sgk1*^{fl/fl} mice fed with high salt diet (HSD) or control diet (*n* = 15). **f**, Quantification of CD4⁺IL-17⁺ or CD4⁺IFN-γ⁺ T cells from the indicated organs of *Cd4*^{Cre}*Sgk1*^{+/+} and *Cd4*^{Cre}*Sgk1*^{fl/fl} mice fed HSD or control diet on day 17 after immunization with MOG₃₅₋₅₅-CFA. **P* < 0.05 (Student's *t*-test). Data are representative of three independent experiments. mLN, mesenteric lymph node. Error bars, s.d.

showed a much milder enhancement of T_H17 cell frequency in the gut, whereas there was no increase in IFN- γ production in any of the mice fed with HSD (Supplementary Fig. 8a, b).

Finally, we studied whether HSD would affect the development of T_H17 and EAE *in vivo*. Wild-type mice fed HSD had more severe EAE than mice fed a normal diet, and this increased severity was dramatically reduced in SGK1-deficient mice (Fig. 4e and Supplementary Fig. 8c,d). We also observed a significantly higher frequency of T_H17 cells in mesenteric lymph nodes and CNS of wild-type mice fed with HSD than in those of SGK1-deficient mice fed with HSD. The percentage of IFN- γ producing T cells in the CNS, but not in the peripheral immune compartments, of wild-type mice was increased in mice fed HSD, suggesting that HSD may increase infiltration but not expansion of IFN- γ ⁺ effector T cells in the target organs (Fig. 4f and Supplementary Fig. 8e). Consistent with our *in vitro* data, we observed elevated IL-17 but not IFN- γ production from CD4⁺ T cells isolated from EAE-immunized wild-type mice fed with HSD and restimulated *in vitro* with MOG_{35–55}, compared to production from cells from EAE mice fed a normal diet (Supplementary Fig. 8f). The data presented here indicate that high sodium intake potentiates T_H17 cell generation *in vivo* in an SGK1-dependent manner and therefore has the potential to increase the risk of promoting autoimmunity.

In conclusion, we used a combination of microarray data analysis, large-scale PPI network analysis and experimental data from several different knockout mice to establish IL-23R-SGK1-Foxo1 as a critical axis in T_H17 stabilization. We show that Foxo1 acts as a repressor of IL-23R expression by binding directly to the *Il23r* promoter and inhibiting ROR γ t-mediated *Il23r* transactivation. Phosphorylation of Foxo1, mediated by SGK1, leads to its deactivation and promotes unopposed ROR γ t-mediated *Il23r* transcription. SGK1 has been studied extensively in the context of NaCl transport^{19,20}. Modest increase of the NaCl concentration induces SGK1 expression in T cells with increased IL-23R expression and T_H17 cell generation *in vitro*. Interestingly, even in un-immunized mice fed with HSD, enhancement of T_H17 differentiation was observed *in vivo* in the gut and gut-associated lymphoid tissue, and this increase in T_H17 cells can be recalled at other peripheral sites after immunization. Although our data suggest an essential role for SGK1 in this process, it is likely that other immune cells and pathways are also influenced by increased salt intake. Furthermore, our results do not exclude additional alternative mechanisms by which an increase in NaCl affects T_H17 cells. Nevertheless, the elevated *in vivo* T_H17 differentiation resulting from HSD raises the important issue of whether increased salt in westernized diets and in processed foods contributes to an increased generation of pathogenic T_H17 cells and for an unprecedented increase in autoimmune diseases.

METHODS SUMMARY

Microarrays and network analysis. For gene-expression analysis Affymetrix microarray chips were used. Data were processed using the GenePattern suite²¹. Differentially expressed genes were detected using fold-change and *t*-test analysis (for *Sgk1*^{−/−} and NaCl-treated T cells) or a consensus of fold-change, the EDGE software²² and a novel sigmoid-based method²³ (for the *Il23r*^{−/−} T_H17 cell time-course data). A command-line version of the ANAT software⁹ was used for network analysis.

***In vitro* T cell differentiation.** Naive T cells were FACS-sorted, stimulated with plate-bound anti-CD3 and anti-CD28 antibodies and the indicated cytokines or NaCl, and cells were analysed by qPCR or flow cytometry at different time points.

Experimental autoimmune encephalomyelitis model. Mice were immunized subcutaneously with MOG_{35–55} in complete Freund's adjuvant (CFA), and heat-inactivated *Mycobacterium tuberculosis* and with intraperitoneal injection of *Bordetella pertussis* toxin.

***In vivo* cell transfer.** Naive T cells were differentiated towards T_H17 cells, then transferred into MOG_{35–55}-CFA-immunized hosts and T cells isolated from various organs were analysed by flow cytometry at 7–12 days after onset of EAE.

Western blot and immunoprecipitation. Differentiated T cells or transfected HEK293T cells were lysed in whole cell extract (WCE) buffer (containing 50 mM Tris buffer, pH 7.5, 100 mM NaCl, 0.1% Triton X-100, 10% v/v glycerol, 1 mM DTT, 1 mM PMSF and protease inhibitors (Sigma)), and lysates were subjected to western blot or immunoprecipitation analysis.

Promoter-activity reporter assay. HEK293T cells were transfected with luciferase reporter constructs and expression vectors, and luciferase expression was determined after 48 h.

Received 30 May 2012; accepted 6 February 2013.

Published online 6 March 2013.

- Korn, T., Bettelli, E., Oukka, M. & Kuchroo, V. K. IL-17 and Th17 Cells. *Annu. Rev. Immunol.* **27**, 485–517 (2009).
- Aggarwal, S., Ghilardi, N., Xie, M. H., de Sauvage, F. J. & Gurney, A. L. Interleukin-23 promotes a distinct CD4 T cell activation state characterized by the production of interleukin-17. *J. Biol. Chem.* **278**, 1910–1914 (2003).
- Zhou, L. *et al.* IL-6 programs T_H-17 cell differentiation by promoting sequential engagement of the IL-21 and IL-23 pathways. *Nature Immunol.* **8**, 967–974 (2007).
- Lang, F. *et al.* (Patho)physiological significance of the serum- and glucocorticoid-inducible kinase isoforms. *Physiol. Rev.* **86**, 1151–1178 (2006).
- Wulff, P. *et al.* Impaired renal Na⁺ retention in the sgk1-knockout mouse. *J. Clin. Invest.* **110**, 1263–1268 (2002).
- Salker, M. S. *et al.* Deregulation of the serum- and glucocorticoid-inducible kinase SGK1 in the endometrium causes reproductive failure. *Nature Med.* **17**, 1509–1513 (2011).
- Zhang, L., Cui, R., Cheng, X. & Du, J. Antiapoptotic effect of serum and glucocorticoid-inducible protein kinase is mediated by novel mechanism activating I κ B kinase. *Cancer Res.* **65**, 457–464 (2005).
- Shelly, C. & Herrera, R. Activation of SGK1 by HGF, Rac1 and integrin-mediated cell adhesion in MDCK cells: PI-3K-dependent and -independent pathways. *J. Cell Sci.* **115**, 1985–1993 (2002).
- Yosef, N. *et al.* ANAT: a tool for constructing and analyzing functional protein networks. *Sci. Signal.* **4**, pl1 (2011).
- Brunet, A. *et al.* Protein kinase SGK mediates survival signals by phosphorylating the forkhead transcription factor FKHRL1 (FOXO3a). *Mol. Cell Biol.* **21**, 952–965 (2001).
- Wei, G. *et al.* Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4⁺ T cells. *Immunity* **30**, 155–167 (2009).
- Di Pietro, N. *et al.* Serum- and glucocorticoid-inducible kinase 1 (SGK1) regulates adipocyte differentiation via forkhead box O1. *Mol. Endocrinol.* **24**, 370–380 (2010).
- Essaghir, A., Dif, N., Marbehant, C. Y., Coffey, P. J. & Demoulin, J. B. The transcription of FOXO genes is stimulated by FOXO3 and repressed by growth factors. *J. Biol. Chem.* **284**, 10334–10342 (2009).
- Ciofani, M. *et al.* A validated regulatory network for Th17 cell specification. *Cell* **151**, 289–303 (2012).
- Berer, K. *et al.* Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination. *Nature* **479**, 538–541 (2011).
- Sczesnak, A. *et al.* The genome of th17 cell-inducing segmented filamentous bacteria reveals extensive auxotrophy and adaptations to the intestinal environment. *Cell Host Microbe* **10**, 260–272 (2011).
- Stegbauer, J. *et al.* Role of the renin-angiotensin system in autoimmune inflammation of the central nervous system. *Proc. Natl Acad. Sci. USA* **106**, 14942–14947 (2009).
- Herrada, A. A. *et al.* Aldosterone promotes autoimmune damage by enhancing Th17-mediated immunity. *J. Immunol.* **184**, 191–202 (2010).
- Diakov, A. & Korbmacher, C. A novel pathway of epithelial sodium channel activation involves a serum- and glucocorticoid-inducible kinase consensus motif in the C terminus of the channel's alpha-subunit. *J. Biol. Chem.* **279**, 38134–38142 (2004).
- Loffing, J. *et al.* Aldosterone induces rapid apical translocation of ENaC in early portion of renal collecting system: possible role of SGK. *Am. J. Physiol. Renal Physiol.* **280**, F675–F682 (2001).
- Reich, M. *et al.* GenePattern 2.0. *Nature Genet.* **38**, 500–501 (2006).
- Storey, J. D., Xiao, W., Leek, J. T., Tompkins, R. G. & Davis, R. W. Significance analysis of time course microarray experiments. *Proc. Natl Acad. Sci. USA* **102**, 12837–12842 (2005).
- Chechik, G. & Koller, D. Timing of gene expression responses to environmental changes. *J. Comput. Biol.* **16**, 279–290 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Kozoriz for cell sorting and A. Waisman for providing *Il17f*^{Cre} mice. L. Zhou, D. Accili, J. Demoulin and K. Sato provided reagents. This work was supported by the US National Institutes of Health (NS030843, NS045937, AI073748 and AI045757 to V.K.K.; 1P01HG005062-01, 1P50HG006193-01 and DP1-OD003958-01 to A.R.; and K01DK090105 to S.X.); the National MS Society (RG2571 to V.K.K.); the Howard Hughes Medical Institute (A.R.); the Klarman Cell Observatory; Guthy Jackson Foundation; and the Austrian Science Fund (FWF, J 3091-B12 to T.T.).

Author Contributions C.W., N.Y. and T.T. carried out experiments and wrote the manuscript. C.Z., S.X. and Y.K. carried out experiments. N.Y. analysed the data. A.R. and V.K.K. supervised the study and edited the manuscript.

Author Information The microarray data sets have been deposited in the Gene Expression Omnibus database under accession numbers GSE43956, GSE43957 and GSE43969. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.R. (aregev@broad.mit.edu) or V.K.K. (vkuchroo@rics.bwh.harvard.edu).

Sodium chloride drives autoimmune disease by the induction of pathogenic T_H17 cells

Markus Kleinewietfeld^{1,2}, Arndt Manzel^{3,4}, Jens Titze^{5,6}, Heda Kvakan^{7,8}, Nir Yosef², Ralf A. Linker³, Dominik N. Müller^{7,9*} & David A. Hafler^{1,2*}

There has been a marked increase in the incidence of autoimmune diseases in the past half-century. Although the underlying genetic basis of this class of diseases has recently been elucidated, implicating predominantly immune-response genes¹, changes in environmental factors must ultimately be driving this increase. The newly identified population of interleukin (IL)-17-producing CD4⁺ helper T cells (T_H17 cells) has a pivotal role in autoimmune diseases². Pathogenic IL-23-dependent T_H17 cells have been shown to be critical for the development of experimental autoimmune encephalomyelitis (EAE), an animal model for multiple sclerosis, and genetic risk factors associated with multiple sclerosis are related to the IL-23–T_H17 pathway^{1,2}. However, little is known about the environmental factors that directly influence T_H17 cells. Here we show that increased salt (sodium chloride, NaCl) concentrations found locally under physiological conditions *in vivo* markedly boost the induction of murine and human T_H17 cells. High-salt conditions activate the p38/MAPK pathway involving nuclear factor of activated T cells 5 (NFAT5; also called TONEBP) and serum/glucocorticoid-regulated kinase 1 (SGK1) during cytokine-induced T_H17 polarization. Gene silencing or chemical inhibition of p38/MAPK, NFAT5 or SGK1 abrogates the high-salt-induced T_H17 cell development. The T_H17 cells generated under high-salt conditions display a highly pathogenic and stable phenotype characterized by the upregulation of the pro-inflammatory cytokines GM-CSF, TNF- α and IL-2. Moreover, mice fed with a high-salt diet develop a more severe form of EAE, in line with augmented central nervous system infiltrating and peripherally induced antigen-specific T_H17 cells. Thus, increased dietary salt intake might represent an environmental risk factor for the development of autoimmune diseases through the induction of pathogenic T_H17 cells.

Although we have recently elucidated many of the genetic variants underlying the risk of developing autoimmune diseases¹, the significant increase in disease incidence, particularly of multiple sclerosis and type 1 diabetes, indicates that there have been fundamental changes in the environment that cannot be related to genetic factors. Diet has long been postulated as a potential environmental risk factor for this increasing incidence of autoimmune diseases in developed countries over recent decades³. One such dietary factor, which rapidly changed along with the Western diet and increased consumption of processed foods or 'fast foods', is salt (NaCl)^{4,5}. The salt content in processed foods can be more than 100 times higher in comparison to similar home-made meals^{5,6}.

We have shown that excess NaCl uptake can affect the innate immune system⁷. Macrophages residing in the skin interstitium modulate local electrolyte composition in response to NaCl-mediated extracellular hypertonicity, and their regulatory activity provides a

buffering mechanism for salt-sensitive hypertension⁷. Moreover, blockade of the renin-angiotensin system, can modulate immune responses and affect EAE^{8,9}. Thus, to investigate whether increased NaCl intake might have a direct effect on CD4⁺ T-cell populations and therefore represents a risk factor for autoimmune diseases, we investigated the effect of NaCl on the *in vitro* differentiation of human T_H17 cells. We induced hypertonicity by increasing NaCl concentration by 10–40 mM (high-salt) in the culture medium and thus mimicked concentrations that could be found in the interstitium of animals fed a high-salt diet⁷. As we previously reported, T_H17-promoting conditions for naive CD4⁺ cells only induced a mild T_H17 phenotype¹⁰. Surprisingly, stimulation under increased NaCl concentrations markedly induced naive CD4⁺ cell expression of IL-17A as determined by flow cytometry

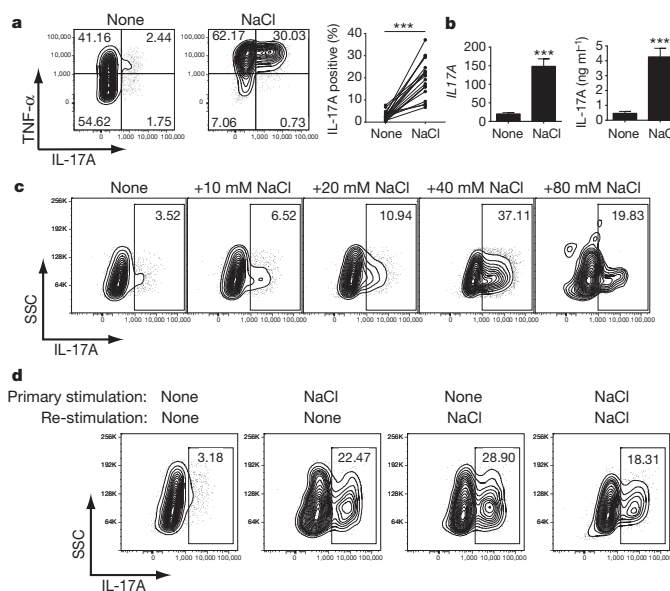


Figure 1 | Sodium chloride promotes the stable induction of T_H17 cells.

a, Naive CD4⁺ cells were differentiated into T_H17 cells in the presence (NaCl) or absence (none) of additional 40 mM NaCl and analysed by flow cytometry (FACS) for IL-17A ($n = 20$). **b**, IL-17A expression was measured by qRT-PCR (left panel, $n = 10$) and ELISA (right panel, $n = 5$). **c**, Cells were stimulated as in **a** under the indicated increased NaCl concentrations and analysed by FACS (one representative experiment of five is shown). **d**, Cells were stimulated as in **a** and were rested in the presence of IL-2. After 1 week, cells were re-stimulated as in **a** in the presence or absence of NaCl for another week and analysed by FACS (one representative experiment of five is shown). *** $P < 0.001$. qRT-PCR data are depicted as relative expression. For all figures, error bars show, unless indicated elsewhere, mean \pm s.e.m.

¹Departments of Neurology and Immunobiology, Yale School of Medicine, 15 York Street, New Haven, Connecticut 06520, USA. ²Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ³Department of Neurology, University of Erlangen-Nuremberg, Schwabachanlage 6, 91054 Erlangen, Germany. ⁴International Graduate School of Neuroscience, Ruhr-University Bochum, Universitätsstr. 150, 44801 Bochum, Germany. ⁵Division of Clinical Pharmacology, Vanderbilt University, 2213 Garland Avenue, Nashville, Tennessee 37232, USA. ⁶Interdisciplinary Center for Clinical Research and Department for Nephrology and Hypertension, University of Erlangen-Nuremberg, Glückstr. 6, 91054 Erlangen, Germany. ⁷Experimental and Clinical Research Center, a joint cooperation between the Charité Medical Faculty and the Max-Delbrück Center for Molecular Medicine, Lindenberger Weg 80, 13125 Berlin, Germany. ⁸Helios Klinikum Berlin-Buch, Schwanebecker Chaussee 50, 13125 Berlin, Germany. ⁹Nikolaus-Fiebiger-Center for Molecular Medicine, University of Erlangen-Nuremberg, Glückstr. 6, 91054 Erlangen, Germany.

*These authors contributed equally to this work.

(Fig. 1a) or by quantitative polymerase chain reaction with reverse transcription (qRT-PCR) and enzyme-linked immunosorbent assay (ELISA) (Fig. 1b). The effect was dose dependent and an optimum of IL-17A induction was achieved by adding 40 mM NaCl in the presence of T_H17 -inducing cytokines (TGF- β 1, IL-1 β , IL-6, IL-21, IL-23) (Fig. 1c and Supplementary Fig. 1). As expected, TNF- α was also induced¹¹, and increasing salt concentrations further led to cell death (data not shown). Nevertheless, adding 40 mM NaCl was tolerated by $CD4^+$ cells with little effect on growth or apoptosis (Supplementary Fig. 2). We then examined whether the nature of cation, anion, or osmolality drives the increases in IL-17A secretion. We found that adding 40 mM sodium gluconate delivered an almost similar degree of T_H17 induction, whereas mannitol or $MgCl_2$ had only a slight effect. Moreover, 80 mM urea, an osmolyte able to pass through cell membranes, had no effect (Supplementary Fig. 3). Thus, the sodium cation was critical for IL-17A induction. We next examined the stability of the salt-induced effect. Naive $CD4^+$ cells that were initially stimulated under high-salt conditions continued to express increased amounts of IL-17A if re-stimulated under normal-salt conditions but could not be further induced with additional salt re-stimulation (Fig. 1d). This is consistent with the observation that only naive but not memory $CD4^+$ cells respond efficiently to increased salt concentrations (Supplementary Fig. 4). The high-salt effect was also observed when T_H17 cells were induced by antigen-specific stimulation (Supplementary Fig. 5)¹². Furthermore, the effect was largely specific for T_H17 cells, as we did not observe comparable outcomes on differentiation of T_H1 or T_H2 cells (Supplementary Fig. 6).

To examine the mechanisms of enhanced IL-17A induction we performed a microarray analysis of naive $CD4^+$ T cells differentiated in the presence or absence of high-salt conditions (Fig. 2a and Supplementary Fig. 8). These data confirmed that cells displayed a stronger T_H17 phenotype under high-salt conditions, as most key signatures of T_H17 cells^{2,13} including *CCL20*, *IL17F*, *RORC* and *IL23R* expression were highly upregulated. The analysis of the microarray data and its verification on messenger RNA or protein expression indicated that high-salt conditions induce a pathogenic type of T_H17 cells¹⁴. In addition to IL-17A, high NaCl concentration induced the expression of pro-inflammatory cytokines IL-2, TNF- α , IL-9 and several chemokines. These cells also upregulated CSF2 (also called GM-CSF), which is essential for the pathogenicity of T_H17 cells^{15,16}, and CCR6, which is crucial for T_H17 function in autoimmune disease¹⁷. Furthermore, *MIR155HG* (also called *MIRHG2*), the host gene for the microRNA miR-155 which is necessary for T_H17 -induced EAE, was highly upregulated¹⁸. The high-salt-induced T_H17 cells also expressed more *TBX21* (also called T-bet) and less *GATA3* and *CXCR6* (Fig. 2a, b and Supplementary Figs 7 and 8, and data not shown). In total, these observations indicate that increased NaCl concentrations specifically promote the generation of a highly pathogenic T_H17 cell type¹⁴.

We then examined the pathways whereby high-salt concentration induced this inflammatory phenotype. It has been shown that increased NaCl concentrations associated with augmented hypertonicity could induce immune system activation^{11,19}. Moreover, it is known that hypertonic stress in mammals is sensed through p38/MAPK, a homologue to HOG1, the ancient yeast hypertonic stress-response element¹⁹. The key translator of this cascade is the osmosensitive transcription factor NFAT5 (refs 20, 21). Analysis of the microarray data set indicated the stimulation of both inflammatory and classic hypertonicity induced pathways. The $CD4^+$ cells expressed high levels of the NFAT5 targets *SGK1* (ref. 22) and the sodium/myo-inositol co-transporter *SLC5A3* (Fig. 2a, b and Supplementary Figs 7 and 8)^{21,23}. Therefore, we proposed that increased NaCl concentration leads to phosphorylation of p38/MAPK that activates other downstream targets, including NFAT5. The phosphorylation of p38/MAPK was indeed increased in the presence of high-salt conditions (Fig. 3a and Supplementary Fig. 9a) and was accompanied by induction of *NFAT5* expression (Fig. 3c). We then determined whether inhibition of the p38/MAPK pathway influenced the effect. SB202190, an inhibitor of p38/MAPK²¹ (p38i), only partially decreased *NFAT5* mRNA induction (Fig. 3c); however, SB202190 sharply reduced T_H17 polarization (Fig. 3b). In line with these findings, short interfering RNA (siRNA)-mediated knockdown of *MAPK14* in $CD4^+$ cells led to less IL-17A production (Supplementary Fig. 9b). High-salt concentration could also promote p38/MAPK activation via the release of ATP²⁴. However, by interfering with this pathway we could not observe significant changes on T_H17 differentiation (data not shown).

Our data indicate that NFAT5 is involved in this NaCl-induced inflammatory pathway. Because it has been shown previously that NFAT5 influences responses of immune cells under similar conditions^{7,20,21}, we silenced *NFAT5* by a short hairpin RNA (shRNA) in naive $CD4^+$ cells. As expected, *NFAT5* silencing reduced *SLC5A3* expression, but also decreased IL-17A and CCR6 expression (Fig. 3d). A direct downstream target of NFAT5 is *SGK1* (ref. 22). Besides being activated by tonicity-dependent signals^{25,26}, *SGK1* expression is also regulated by TGF- β ²⁷ and glucocorticoids²⁸. As *SGK1* activation can be regulated by p38/MAPK^{26,29} and NFAT5²², and was strongly upregulated in the microarray, it was of interest to examine whether this kinase has a role in high-salt-mediated T_H17 polarization. *SGK1* was upregulated in naive $CD4^+$ cells after stimulation with high-salt conditions. To confirm that *SGK1* is regulated by p38/MAPK-dependent signals, expression of *SGK1* was measured in the presence of SB202190 (Fig. 3e). The addition of p38i reduced NaCl-induced *SGK1* mRNA expression, consistent with previous reports in other systems^{22,26,29}. Moreover, shRNA-mediated silencing of *SGK1* significantly decreased IL-17A production in high-salt exposed cells and led to diminished CCR6 expression (Fig. 3f). In line with these observations, pharmacological blockade

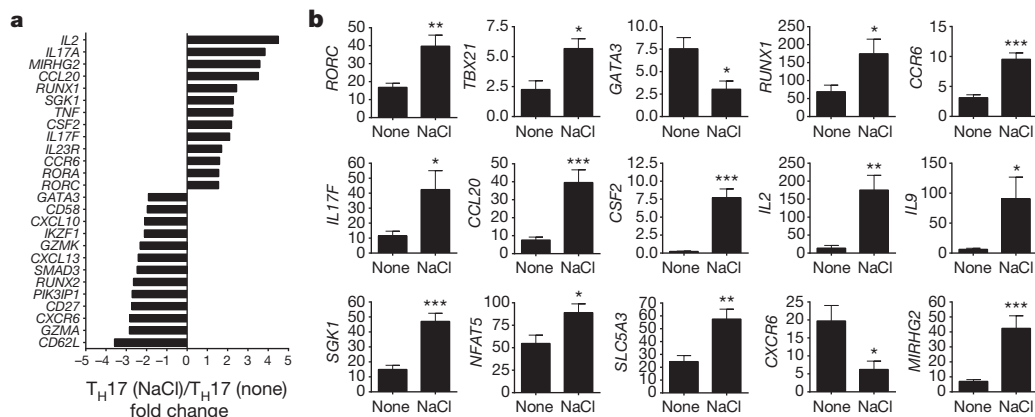


Figure 2 | High-salt-induced T_H17 cells display a pathogenic phenotype. **a**, Microarray analysis of naive $CD4^+$ cells differentiated into T_H17 cells in the presence (NaCl) or absence (none) of additional 40 mM NaCl. Depicted is a

selection of 26 up- and downregulated genes (mean fold change of two independent experiments). **b**, qRT-PCR analysis of differentially expressed genes in the two groups ($n = 5-8$). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

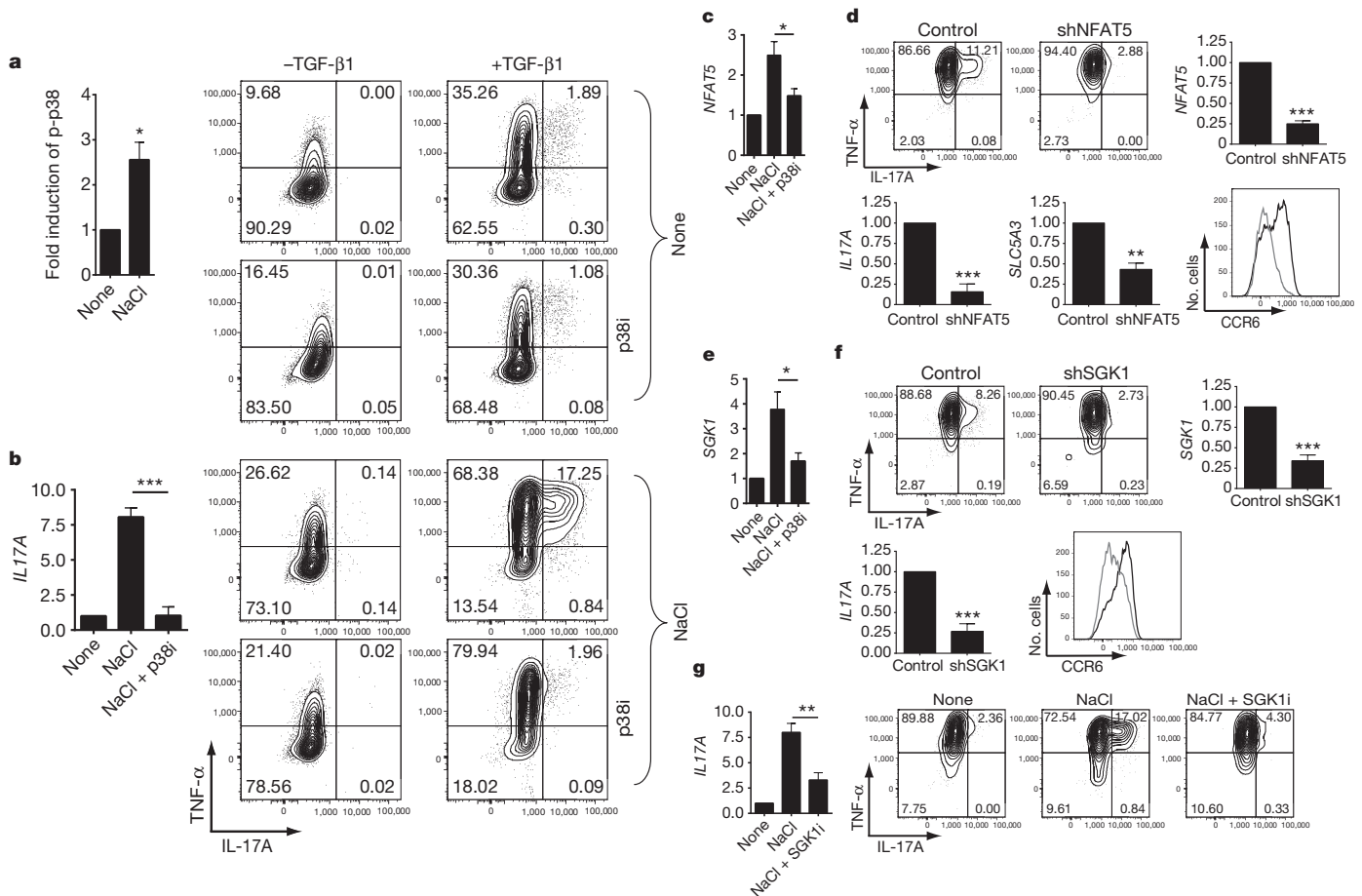


Figure 3 | The induction of T_H17 cells by NaCl depends on p38/MAPK, NFAT5 and SGK1. **a**, Naive CD4⁺ cells were stimulated in the presence (NaCl) or absence (none) of additional 40 mM NaCl and were analysed by FACS for phosphorylated p38 (p-p38; *n* = 5). **b**, Naive CD4⁺ cells were differentiated into T_H17 cells as indicated in the presence or absence of NaCl and SB202190 (p38i) and analysed by qRT-PCR as depicted in the bar graph (*n* = 7) or by FACS (the left row shows cells differentiated in the absence of TGF-β1). **c**, Naive CD4⁺ cells were stimulated for 3 h in the presence or absence of NaCl and SB202190 and analysed by qRT-PCR for *NFAT5* (*n* = 4). **d**, Cells were transduced with NFAT5-specific (shNFAT5) or control shRNA (control), stimulated as in **b** and analysed by FACS. The bar graphs depict qRT-PCR analyses of *NFAT5*, *IL17A* and *SLC5A3* (*n* = 5). CCR6 was analysed by FACS (black histogram, control; grey histogram, shNFAT5; displayed as cell number versus CCR6; one representative experiment of four is shown). **e**, Cells were stimulated as in **c** but analysed by qRT-PCR for *SGK1* (*n* = 4). **f**, Cells were transduced with a shRNA specific for *SGK1* (shSGK1) or a control shRNA (control) and activated as in **b**, and analysed by FACS. Expression of *SGK1* and *IL17A* was determined by qRT-PCR (*n* = 5). CCR6 was analysed by FACS (black histogram, control; grey histogram, shSGK1; displayed as cell number versus CCR6; one representative experiment of four is shown). **g**, Cells were cultured as in **b** but in the presence or absence of the SGK1 inhibitor GSK650394 (SGK1i) and analysed by FACS. The bar graph shows qRT-PCR for *IL17A* under similar conditions (*n* = 5). FACS and qRT-PCR (relative expression) data depicted in bar graphs were normalized to controls. **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

of SGK1 produced similar, albeit less pronounced, results compared to SB202190 (Fig. 3g).

The rather dramatic *in vitro* effects of high-salt concentration on naive human CD4⁺ cells prompted us to examine the effects of increased dietary NaCl in an *in vivo* system. We first adapted the human culture system to various murine T_H17 differentiation models and made similar observations of increased T_H17 induction and the accompanying phenotype (Fig. 4 and Supplementary Fig. 10). High-salt conditions did not significantly alter proliferation or cell death. Moreover, the effect was specific for T_H17 conditions, as there was no enhancement of T_H1 or T_H2 differentiations (Supplementary Figs 11 and 12 and data not shown). High-salt-induced expression of *Nfat5*, *Sgk1* and *IL-17A* was dependent on p38/MAPK. Enhanced T_H17 differentiation could be blocked by SB202190, and gene deletion of p38α decreased *Il17a*, *Nfat5* and *Sgk1* induction (Supplementary Figs 10 and 13). As the high-salt effect on T_H17 cells appeared similar between species, we examined whether dietary NaCl influenced EAE. High-salt diet accelerated onset and increased severity of the disease (Fig. 4c), whereas blood pressure was not affected (Supplementary Fig. 14). Mice on the high-salt diet displayed significantly higher numbers of CNS-infiltrating CD3⁺ and Mac3⁺ cells compared to controls (Fig. 4c).

IL-17A-expressing CD4⁺ cells in CNS infiltrates almost doubled in frequency and, accordingly, we detected increased *Il17a* and *Rorc* mRNA expression in spinal cords (Fig. 4d and Supplementary Fig. 15). In contrast to *Ifng*, we found augmented expression of *Il17a* and *Csf2* and higher levels of *Nfat5* and *Sgk1* in the spleens of high-salt diet EAE mice compared to controls (Fig. 4e). Notably, splenocytes from EAE mice fed the high-salt diet showed enhanced IL-17A but not IFN-γ or T_H2 cytokine expression upon antigen re-stimulation, indicating increased *in vivo* induction of antigen-specific T_H17 cells (Fig. 4f and data not shown). Consistent with *in vitro* data, the high-salt-diet-induced effect was dependent on p38/MAPK, as *in vivo* administration of SB202190 inhibited salt-induced increases in the frequency of T_H17 cells infiltrating the CNS (Supplementary Fig. 15b, c).

In this investigation, we found that modest increases in NaCl concentration could stimulate an almost logarithmic *in vitro* induction of IL-17A in naive CD4⁺ cells mediated through p38/MAPK, NFAT5 and SGK1. Importantly, the addition of 40 mM of NaCl to T_H17 differentiation cultures not only increased IL-17A expression but also led to a pathogenic phenotype of T_H17 cells. In line with these findings, common salt added to the diet of mice led to severe worsening of EAE accompanied by increased numbers of T_H17 cells.

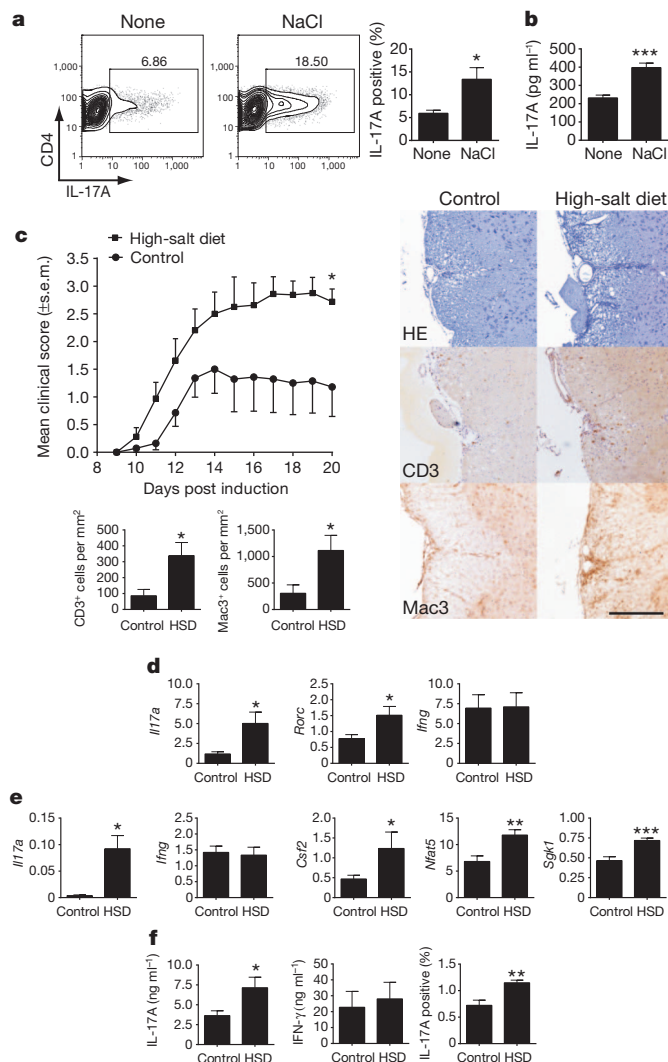


Figure 4 | High-salt diet induces TH17 cells *in vivo* and exacerbates experimental autoimmune encephalomyelitis. **a**, Naive murine CD4⁺ cells were stimulated with irradiated APCs, anti-CD3, IL-6 and TGF- β 1 in the presence (NaCl) or absence (none) of additional 40 mM NaCl and were analysed by FACS ($n = 3$). **b**, IL-17A secretion (ELISA) of primary splenocytes, stimulated by anti-CD3 in the presence or absence of NaCl ($n = 6$). **c**, Mean clinical scores of EAE in high-salt diet (HSD) animals (squares) or controls (dots, pooled data of two independent experiments with 12 animals). Histological analyses (right) show sections of the spinal cord stained with haematoxylin and eosin (HE), anti-CD3 and anti-Mac3 for control or HSD animals (scale bar, 100 μ m) and were quantified for CD3 and Mac3 (bar graphs, $n = 5-6$). **d**, Spinal cord from EAE animals was analysed by qRT-PCR ($n = 5-6$). **e**, Splenocytes from EAE animals were analysed by qRT-PCR ($n = 4-7$). **f**, Splenocytes from EAE animals were re-stimulated with MOG for 2 days and supernatants were analysed for IL-17A and IFN- γ by ELISA ($n = 7-8$) or cells were analysed for IL-17A by FACS ($n = 4$). qRT-PCR data are depicted as relative expression. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

What might be the physiological role for the effect of high-salt on the induction of inflammatory TH17 cells? The concentration of Na⁺ in plasma is approximately 140 mM, similar to standard cell culture media. Less well appreciated is that in the interstitium and lymphoid tissue, considerably higher Na⁺ concentrations between 160 mM and even as high as 250 mM can be encountered^{7,20}—the ‘high-salt’ conditions that we found to induce inflammatory TH17 cells. Thus, this may be a mechanism for decreasing immune activation in the blood while favouring an inflammatory response in lymphoid tissues or with migration of cells into tissue. In this context it could be expected that other immune cells can react on high-salt conditions as well and potentially contribute to the effects observed *in vivo*.

Do these data indicate that increased salt intake is the long-sought-after environmental factor associated with the epidemic of autoimmune disease? Although these data present an attractive hypothesis, the direct causality of salt intake and incidence of autoimmune disease is yet to be demonstrated. That is, no *in vitro* observation can prove causality in humans; instead, our data indicate that clinical trials with severe curtailment of salt intake for individuals at risk for developing autoimmune disease are required. Clinical scenarios in which a dietary salt restriction protocol could be tested are multiple sclerosis or psoriasis, both autoimmune diseases with strong TH17 components². Additionally, excess salt content in diet should be investigated as a potential environmental risk factor for autoimmune diseases. However, this study would be difficult in Western cultures where the application of a true low-salt diet, representing the conditions in which *Homo sapiens* were environmentally selected in Africa, is difficult to achieve. Nevertheless, although there might be additional mechanisms contributing to the observed effects, the pathways identified in this study may offer new targets for the treatment of autoimmune diseases, with interference in the p38/MAPK, NFAT5 and SGK1 pathways aimed at blocking the generation of pathogenic TH17 cells.

METHODS SUMMARY

Human cell sorting. Peripheral blood mononuclear cells (PBMCs) were obtained from the peripheral blood of healthy subjects in compliance with institutional review board (IRB) protocols. CD4⁺ T cells were isolated by negative selection using magnetic beads (Miltenyi Biotec). Subsequently, naive T cell were sorted as CD4⁺CD25[−]CD127⁺CD45RO[−]CD45RA⁺ and memory cells were obtained by sorting for CD4⁺CD25[−]CD127⁺CD45RO⁺CD45RA[−] on a FACS Aria (BD Biosciences).

Human differentiation assays. Naive, memory or total CD4⁺ T cells were stimulated by plate-bound anti-CD3 and soluble anti-CD28 in serum-free X-VIVO15 medium (BioWhittaker) where indicated in the presence of various cytokines (IL-1 β , IL-6, IL-21, IL-23, TGF- β 1) and different concentrations of NaCl. Cells were analysed for cytokine expression by intracellular flow cytometry. Cytokine secretion was measured by ELISA (eBioscience). mRNA expression was determined by quantitative RT-PCR (Applied Biosystems).

EAE induction and high-salt diet. Male C57BL/6J mice (Harlan) were immunized with 200 μ g MOG₃₅₋₅₅ in an equal amount of complete Freund's adjuvant and received 200 ng pertussis toxin intraperitoneally on days 0 and 2 post induction. The clinical evaluation was performed daily on a 5 point scale ranging from 0 (no clinical sign) to 5 (moribund). Mice received normal chow and tap water ad libitum (control) or sodium-rich chow containing 4% NaCl and tap water containing 1% NaCl ad libitum (high-salt diet).

Full Methods and any associated references are available in the online version of the paper.

Received 13 March; accepted 19 December 2012.

Published online 6 March 2013.

1. International Multiple Sclerosis Genetics Consortium & Wellcome Trust Case Control Consortium. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).
2. Korn, T., Bettelli, E., Oukka, M. & Kuchroo, V. K. IL-17 and Th17 Cells. *Annu. Rev. Immunol.* **27**, 485–517 (2009).
3. Ascherio, A. & Munger, K. L. Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors. *Ann. Neurol.* **61**, 504–513 (2007).
4. McGuire, S. Institute of Medicine. 2010. Strategies to reduce sodium intake in the United States. Washington, DC: The National Academies Press. *Adv. Nutr.* **1**, 49–50 (2010).
5. Appel, L. J. *et al.* The importance of population-wide sodium reduction as a means to prevent cardiovascular disease and stroke: a call to action from the American Heart Association. *Circulation* **123**, 1138–1143 (2011).
6. Brown, I. J., Tzoulaki, I., Candeias, V. & Elliott, P. Salt intakes around the world: implications for public health. *Int. J. Epidemiol.* **38**, 791–813 (2009).
7. Machnik, A. *et al.* Macrophages regulate salt-dependent volume and blood pressure by a vascular endothelial growth factor-C-dependent buffering mechanism. *Nature Med.* **15**, 545–552 (2009).
8. Platten, M. *et al.* Blocking angiotensin-converting enzyme induces potent regulatory T cells and modulates TH1- and TH17-mediated autoimmunity. *Proc. Natl Acad. Sci. USA* **106**, 14948–14953 (2009).
9. Stegbauer, J. *et al.* Role of the renin-angiotensin system in autoimmune inflammation of the central nervous system. *Proc. Natl Acad. Sci. USA* **106**, 14942–14947 (2009).

10. Yang, L. *et al.* IL-21 and TGF- β are required for differentiation of human T_H17 cells. *Nature* **454**, 350–352 (2008).
11. Junger, W. G., Liu, F. C., Loomis, W. H. & Hoyt, D. B. Hypertonic saline enhances cellular immune function. *Circ. Shock* **42**, 190–196 (1994).
12. Zielinski, C. E. *et al.* Pathogen-induced human TH17 cells produce IFN- γ or IL-10 and are regulated by IL-1 β . *Nature* **484**, 514–518 (2012).
13. Zhou, L. & Littman, D. R. Transcriptional regulatory networks in Th17 cell differentiation. *Curr. Opin. Immunol.* **21**, 146–152 (2009).
14. Ghoreschi, K. *et al.* Generation of pathogenic T_H17 cells in the absence of TGF- β signalling. *Nature* **467**, 967–971 (2010).
15. Codarri, L. *et al.* ROR γ t drives production of the cytokine GM-CSF in helper T cells, which is essential for the effector phase of autoimmune neuroinflammation. *Nature Immunol.* **12**, 560–567 (2011).
16. El-Behi, M. *et al.* The encephalitogenicity of T_H17 cells is dependent on IL-1- and IL-23-induced production of the cytokine GM-CSF. *Nature Immunol.* **12**, 568–575 (2011).
17. Reboldi, A. *et al.* C-C chemokine receptor 6-regulated entry of T_H17 cells into the CNS through the choroid plexus is required for the initiation of EAE. *Nature Immunol.* **10**, 514–523 (2009).
18. O'Connell, R. M. *et al.* MicroRNA-155 promotes autoimmune inflammation by enhancing inflammatory T cell development. *Immunity* **33**, 607–619 (2010).
19. Shapiro, L. & Dinarello, C. A. Osmotic regulation of cytokine synthesis *in vitro*. *Proc. Natl Acad. Sci. USA* **92**, 12230–12234 (1995).
20. Go, W. Y., Liu, X., Roti, M. A., Liu, F. & Ho, S. N. NFAT5/TonEBP mutant mice define osmotic stress as a critical feature of the lymphoid microenvironment. *Proc. Natl Acad. Sci. USA* **101**, 10673–10678 (2004).
21. Kino, T. *et al.* Brx mediates the response of lymphocytes to osmotic stress through the activation of NFAT5. *Sci. Signal.* **2**, ra5 (2009).
22. Chen, S. *et al.* Tonicity-dependent induction of Sgk1 expression has a potential role in dehydration-induced natriuresis in rodents. *J. Clin. Invest.* **119**, 1647–1658 (2009).
23. Ortells, M. C. *et al.* Transcriptional regulation of gene expression during osmotic stress responses by the mammalian target of rapamycin. *Nucleic Acids Res.* **40**, 4368–4384 (2012).
24. Woehrle, T. *et al.* Hypertonic stress regulates T cell function via pannexin-1 hemichannels and P2X receptors. *J. Leukoc. Biol.* **88**, 1181–1189 (2010).
25. Waldegger, S., Barth, P., Raber, G. & Lang, F. Cloning and characterization of a putative human serine/threonine protein kinase transcriptionally modified during anisotonic and isotonic alterations of cell volume. *Proc. Natl Acad. Sci. USA* **94**, 4440–4445 (1997).
26. Bell, L. M. *et al.* Hyperosmotic stress stimulates promoter activity and regulates cellular utilization of the serum- and glucocorticoid-inducible protein kinase (Sgk) by a p38 MAPK-dependent pathway. *J. Biol. Chem.* **275**, 25262–25272 (2000).
27. Waldegger, S. *et al.* h-sgk serine-threonine protein kinase gene as transcriptional target of transforming growth factor β in human intestine. *Gastroenterology* **116**, 1081–1088 (1999).
28. Webster, M. K., Goya, L., Ge, Y., Maiyar, A. C. & Firestone, G. L. Characterization of sgk, a novel member of the serine/threonine protein kinase gene family which is transcriptionally induced by glucocorticoids and serum. *Mol. Cell. Biol.* **13**, 2031–2040 (1993).
29. Waldegger, S., Gabrysch, S., Barth, P., Fillon, S. & Lang, F. h-sgk serine-threonine protein kinase as transcriptional target of p38/MAP kinase pathway in HepG2 human hepatoma cells. *Cell. Physiol. Biochem.* **10**, 203–208 (2000).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors would like to thank S. Bhela, M. Zaidi, P. Quass, M. Mroz and S. Seubert for technical assistance and F. C. Luft for critical reading of the manuscript. We are grateful to J.-P. David and S. Teufel for providing Mx-Cre⁺ p38^{fl/fl} mice. This work was supported by a National MS Society Collaborative Research Center Award CA1061-A-18, National Institutes of Health Grants P01 AI045757, U19 AI046130, U19 AI070352, and P01 AI039671, and by a Jacob Javits Merit award (NS2427) from the National Institute of Neurological Disorders and Stroke, the Penates Foundation and the Nancy Taylor Foundation for Chronic Diseases, Inc. (to D.A.H.). R.A.L. was supported by the ELAN programme, University of Erlangen. D.N.M. was supported by the German Research Foundation (DFG) and the German Center for Cardiovascular Research (DZHK). J.T. was supported by the Interdisciplinary Center for Clinical Research at University of Erlangen and the German Research Foundation.

Author Contributions M.K. designed the study, planned and performed experiments, analysed data and wrote the manuscript. A.M. planned and performed experiments, analysed data and wrote the manuscript. J.T. and H.K. interpreted data and supported the work with key suggestions and editing the manuscript. N.Y. analysed data. R.A.L. planned experiments, analysed data and wrote the manuscript. D.N.M. designed the study, planned experiments, analysed data and wrote the manuscript. D.A.H. designed the study, planned experiments, analysed data, and wrote the manuscript. M.K., D.N.M. and D.A.H. co-directed the project.

Author Information The microarray data sets have been deposited in the Gene Expression Omnibus database under accession number GSE42569. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.K. (markus.kleinewietfeld@yale.edu) or D.A.H. (david.hafler@yale.edu).

METHODS

Antibodies, recombinant cytokines and reagents. The following monoclonal antibodies and reagents were used as follows: for surface staining, anti-CD4 (RPA-T4), anti-CD45RO (UCHL1), anti-CD45RA (HI100), anti-CD25 (M-A251), anti-CD127 (hIL-7R-M21), anti-CCR6 (11A9) and Annexin V all from BD Biosciences; for intracellular staining, anti-IL-17A (eBio64DEC17), anti-TNF- α (MAB11), anti-IFN- γ (4S.B3), anti-IL-2 (MQ1-17H12), anti-RORC (AFKJS-9), anti-GATA3 (TWAJ) and anti-Tbet (eBio4B10) from eBioscience, and anti-pp38 (36/p38) (BD Biosciences) and anti-GM-CSF (BVD2-21C11) from Biologend; for T-cell stimulation, anti-CD3 (UCHT1) and anti-CD28 (28.2) from BD Biosciences. Recombinant human TGF- β 1 was purchased from eBioscience, recombinant human IL-1 β , IL-6, IL-4, IL-12 and IL-23 and neutralizing anti-IFN- γ (25718) and anti-IL-4 (3007) were purchased from R&D Systems, and recombinant human IL-21 was purchased from Cell Sciences. CFSE was obtained from Invitrogen.

Human cell isolation and stimulation. Peripheral blood was obtained from healthy control volunteers in compliance with Institutional Review Board protocols. Peripheral blood mononuclear cells (PBMCs) were separated by Ficoll-Paque PLUS (GE Healthcare) gradient centrifugation. Untouched total CD4⁺ T cells were isolated from PBMCs by negative selection via the CD4⁺ T-cell isolation kit II (Miltenyi Biotec). Naive (CD45RA⁺CD45RO⁻CD25⁻CD127⁺) and memory (CD45RA⁻CD45RO⁺CD25⁺CD127⁺) CD4⁺ T cells were sorted by high-speed flow cytometry with a FACS Aria (BD Biosciences) to a purity >98% as verified by post-sort analysis. Dead cells were excluded by propidium iodide (BD Biosciences). CD14⁺ monocytes were isolated by positive selection with CD14 microbeads (Miltenyi Biotec). Cells were cultured in 96-well round-bottom plates (Costar) at 5×10^4 cells per well in serum-free X-VIVO15 medium (BioWhittaker), and stimulated with plate-bound anti-CD3 ($10 \mu\text{g ml}^{-1}$) and soluble anti-CD28 ($1 \mu\text{g ml}^{-1}$) antibodies. Where indicated, recombinant TGF- β 1 (5 ng ml^{-1}), IL-1 β (12.5 ng ml^{-1}), IL-6 (25 ng ml^{-1}), IL-21 (25 ng ml^{-1}), or IL-23 (25 ng ml^{-1}) or additional 10–80 mM NaCl was added to the cultures. For T_H1 and T_H2 differentiation, naive cells were stimulated as described above but in the presence of IL-12 (10 ng ml^{-1}) and anti-IL-4 ($10 \mu\text{g ml}^{-1}$) (for T_H1 cells) or with IL-4 (10 ng ml^{-1}) and anti-IFN- γ ($10 \mu\text{g ml}^{-1}$) (for T_H2 cells). In some experiments the specific inhibitors SB202190 (Sigma Aldrich) or GSK650394³⁰ (Tocris Bioscience/R&D Systems) at concentrations of 5 μM or 1 μM , respectively, were added to the cultures. Co-cultures of CD14⁺ monocytes and T cells were performed as described before¹². In brief, monocytes were pulsed for 3 h with *Candida albicans* (GREER) and irradiated (45 Gy) before T cell co-culture (ratio of 1:2). Total CD4⁺ T cells were co-cultured for 12 days and naive CD4⁺ T cells were co-cultured for 7 days in the presence of additional cytokine cocktail (TGF- β 1, IL-1 β , IL-6, IL-21 and IL-23). Recombinant human IL-2 was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH) and was used for re-stimulation experiments at 20 U ml^{-1} . Cells were cultured for the indicated periods of time.

Flow cytometry and cytokine detection. Cells were analysed by flow cytometry if not specified elsewhere after a culture period between 7 and 8 days. For surface staining, cells were stained with the respective antibodies for 20 min in PBS containing 0.5% FCS and 2 mM EDTA before analysis. For intracellular staining, cells were stimulated for 4–5 h with PMA (50 ng ml^{-1}) and ionomycin (250 ng ml^{-1} ; both from Sigma-Aldrich) in the presence of GolgiPlug (BD Biosciences), fixed and made permeable (Fix/Perm; eBioscience) according to the manufacturer's instructions, and stained with the respective antibodies for intracellular cytokine detection for 30–45 min. Before fixation, cells were stained with the LIVE/DEAD cell kit (Invitrogen) to exclude dead cells. For measurement of phosphorylated p38 (p-p38), cells were stimulated for 20 min before cells were fixed (Cytofix buffer, BD Biosciences) and made permeable (Phosflow Perm Buffer III, BD Biosciences) according to the manufacturer's instructions and stained for 30–45 min with anti-p-p38. Data were acquired on a LSR II (BD Biosciences) and analysed with FlowJo software (TreeStar). Culture supernatants were taken on day 6 and measured by ELISA for secretion of IL-17A (eBioscience) according to the manufacturer's instructions.

Real-time PCR. Cells for RNA isolation were harvested if not specified elsewhere between 6 and 7 days of culture and RNA was isolated using the Absolutely RNA 96 Microprep kit (Agilent Technologies) or RNeasy micro kit (Qiagen) and converted to cDNA via reverse transcriptase by random hexamers and Multiscribe RT (TaqMan Gold RT-PCR kit, Applied Biosystems). All primers were purchased from Applied Biosystems. All reactions were performed on a StepOnePlus Real-Time PCR System (Applied Biosystems). The values are represented as the difference in Ct values normalized to β 2-microglobulin for each sample as per the following formula: relative RNA expression = $(2^{-\Delta\text{Ct}}) \times 10^3$.

shRNA- and siRNA-mediated gene silencing. Lentiviral particles expressing shRNAs were obtained from the library of The RNAi Consortium (TRC)³¹.

Lentiviral transduction of human T cells was carried out as described before³². In brief, 5×10^4 human naive CD4⁺ T cells per well were stimulated for 24 h before infection. Cells were then transduced with viral particles containing a vector expressing the indicated specific shRNA or as controls a vector expressing an unspecific shRNA or expressing GFP. Transduction was mediated at a multiplicity of infection (MOI) of 5 by centrifugation at 2,250 r.p.m. for 30 min at room temperature in the presence of $3 \mu\text{g ml}^{-1}$ polybrene (Millipore). After 48 h puromycin (Invitrogen) was added to the cultures at a concentration of $0.5 \mu\text{g ml}^{-1}$ to select for successfully transduced cells and was controlled by flow cytometry for GFP and propidium iodide. The specific RNAi Consortium clones were TRCN0000020019 for NFAT5 and TRCN0000040175 for SGK1. For siRNA transfections, control siRNA (ON-TARGETplus non targeting 1) and a pool of four specific siRNAs for MAPK14 (ON-TARGETplus SMARTpool 1432) were obtained from Thermo Scientific Dharmacon. Cells were transfected by using Human T Cell Nucleofector kit and a Nucleofector II device as recommended by the manufacturer (Lonza/Amara).

Microarray analysis. Cells for microarray analysis were collected at day 7 of culture and total RNA was isolated using Trizol reagent according to the manufacturer (Invitrogen). Expression data were generated by using GeneChip Human Genome U133 Plus 2.0 arrays (Affymetrix) at the Yale Center for Genome Analysis (YCGA). For analysis, the data were normalized using the GenePattern software³³ with the Robust Multi Array (RMA) algorithm³⁴. The COMBAT software was used to remove batch effects. Fold change was computed between the average expression levels of each probe set in samples with the different conditions. To avoid spurious fold levels due to low expression values, a small constant ($c = 50$) was added to the expression values. Only cases where more than 50% of the four possible pair-wise comparisons were over a cutoff of 1.5-fold change were reported. A Z-score was computed as additional filter by comparing the mean of the expression levels in the NaCl-treated samples to the expression levels in the control samples. Only cases with a corresponding *P*-value lower than 0.05 were reported.

Western blotting. Western blotting was performed as described before³⁵. Phospho-p38 was detected by using anti-phospho-p38 (Cell Signaling Technology). Anti- β -actin and anti-SGK1 antibodies were obtained from Cell Signaling Technology and anti-NFAT5 antibodies were purchased from Pierce/Thermo Scientific. Primary antibodies were detected by peroxidase-conjugated streptavidin (Jackson Immuno Research), secondary anti-rabbit-HRP-conjugated (Cell signalling Technology or Jackson Immuno Research) and secondary anti-mouse-HRP-conjugated (Bio-Rad) antibodies.

Mice, EAE induction, high-salt diet and blood pressure analysis. C57BL/6J mice were purchased from Harlan and housed at the in-house animal care facility of the University of Erlangen under standardized conditions. EAE induction was done as described before⁹. Briefly, male mice were immunized with $200 \mu\text{g}$ MOG_{35–55} (Charite) in an equal amount of complete Freund's adjuvant and received 200 ng pertussis toxin (List Biochemicals) intraperitoneally on days 0 and 2 post induction. The clinical evaluation was performed on a daily bases by a 5-point scale ranging from 0, no clinical sign; 1, limp tail; 2, limp tail, impaired righting reflex, and paresis of one limb; 3, hindlimb paralysis; 4, hindlimb and forelimb paralysis; 5, moribund. Mice received normal chow and tap water ad libitum (control group) or sodium-rich chow containing 4% NaCl (SSNIFF) and tap water containing 1% NaCl ad libitum (high-salt group). Inhibition of p38/MAPK *in vivo* was done as described before³⁶. In brief, mice were maintained on a control or high-salt diet and either received $1 \text{ mg kg}^{-1} \text{ d}^{-1}$ SB202190 (TOCRIS) intraperitoneally or vehicle from day –3 post induction of EAE. Brain leukocytes were isolated by percoll gradient centrifugation on day 17 post EAE induction, stimulated by PMA/ionomycin and analysed by flow cytometry for IL-17A and CD4 expression. Mx-Cre⁺/p38 $\alpha^{\text{fl/fl}}$ mice³⁷ maintained on a C57BL/6 background were a gift from J.-P. David. Mice were injected with 13 mg kg^{-1} body weight polyinosinic-polycyidylic acid (poly(I:C), Sigma-Aldrich) on days 0, 2, 6 and were killed on day 8 for isolation of splenocytes. Blood pressure analysis was performed by the tail cuff method as described previously⁹. All animal experimentation was performed in accordance to the German animal protection law.

Histology. On day 20 post induction, mice were perfused with 4% paraformaldehyde and then the lumbar, thoracic and cervical part of their spinal cord was embedded in paraffin. Spinal cord cross-sections were stained with haematoxylin and eosin to assess inflammation. T cells were labelled by anti-CD3 (Serotec), macrophages/microglia by anti-Mac3 (BD Biosciences) and IL-17-positive cells by anti-IL-17 (Abcam).

Murine T-cell cultures. Splenic T cells from EAE animals were re-stimulated with $20 \mu\text{g ml}^{-1}$ MOG_{35–55} peptide for 48 h, for intracellular IL-17A detection, monensin (BD Biosciences) was added to the cultures for an additional 6 h. Splenic T cells from naive mice were stimulated with $1 \mu\text{g ml}^{-1}$ plate-bound anti-CD3 (17A2, BD Biosciences) and $1 \mu\text{g ml}^{-1}$ soluble anti-CD28 (37.51, BD

Biosciences) for 48 h. For T_H17 cell differentiation, spleen and lymph node cells from 10-week-old 2D2 mice³⁸ were pooled and $CD4^+CD62L^+$ naive T cells were isolated by magnetic cell sorting (Miltenyi Biotec). Cells were cultured at 2×10^6 cells ml^{-1} and stimulated for 4 days with 2×10^7 irradiated (30 Gy) syngenic splenocytes per ml and $1 \mu g\ ml^{-1}$ anti-CD3 (2C11, BD Biosciences) in the presence of TGF- $\beta 1$ ($5\ ng\ ml^{-1}$) and IL-6 ($20\ ng\ ml^{-1}$) and where indicated of additional $40\ mM$ NaCl. For APC free T_H17 differentiations, naive T cells were sorted as $CD4^+CD62L^+CD44^{lo}CD25^-$ and stimulated by plate-bound anti-CD3 ($2\ \mu g\ ml^{-1}$) and anti-CD28 ($2\ \mu g\ ml^{-1}$) in the presence of IL-6 ($40\ ng\ ml^{-1}$) and TGF- $\beta 1$ ($1\ ng\ ml^{-1}$) or IL-6 ($40\ ng\ ml^{-1}$) and IL-23 ($10\ ng\ ml^{-1}$) (all from R&D Systems) and were cultured for 4 days. In some experiments, $10\ \mu M$ SB202190 (TOCRIS) was added to the cultures. For T_H1 differentiation, naive $CD4^+$ T cells were cultured for 96 h with anti-CD3, anti-CD28, IL-12 ($20\ ng\ ml^{-1}$) (BioLegend) and anti-IL-4 ($10\ \mu g\ ml^{-1}$) (1B11, BioLegend). To monitor proliferation, cells were labelled with fixable proliferation dye (eBioscience) according to the manufacturer's protocol. For intracellular flow cytometry, cells were stimulated for 4 h with PMA/ionomycin in the presence of monensin and stained for CD4 (RM4-5, eBioscience) and intracellular IL-17A (eBio17B7, eBioscience), IFN- γ (XMG1.2, eBioscience), Tbet (4B10, eBioscience) or RORC/ROR γt (AFKJS-9, eBioscience), excluding dead cells by a fixable viability dye (eBioscience). For murine gene expression analysis, mRNA was prepared using PeqLab Gold HP total RNA kit (PeqLab) and cDNA was prepared using superscript II reverse transcriptase (Invitrogen). RNA was isolated from EAE animals at day 14 post induction. Reactions were performed on a 7900 Sequence Detection System (Applied Biosystems). Primers were obtained from Applied Biosystems and target expression was normalized to β -actin expression. For cytokine secretion analysis, cells were stimulated as indicated and supernatants were collected after 3 days of

culture. Monoclonal antibody pairs and recombinant cytokine standards were purchased from R&D systems (IL-17A, IFN- γ).

Statistical analysis. Statistical analysis was performed using GraphPad Prism (GraphPad Software). Data were analysed by an unpaired *t*-test in case of two groups and by one-way ANOVA using Tukey's post-hoc test in multiple groups. Data tested against a specified value were analysed by a one-sample *t*-test. EAE was analysed using a non-parametric Mann-Whitney *U*-test. Data were presented if not indicated elsewhere as mean \pm s.e.m. $P < 0.05$ was considered to be statistically significant (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

30. Sher, A. B. *et al.* Development of a small-molecule serum- and glucocorticoid-regulated kinase-1 antagonist and its evaluation as a prostate cancer therapeutic. *Cancer Res.* **68**, 7475–7483 (2008).
31. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283–1298 (2006).
32. Astier, A. L. *et al.* RNA interference screen in primary human T cells reveals FLT3 as a modulator of IL-10 levels. *J. Immunol.* **184**, 685–693 (2010).
33. Reich, M. *et al.* GenePattern 2.0. *Nature Genet.* **38**, 500–501 (2006).
34. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
35. Wenzel, K. *et al.* Potential relevance of $\alpha 1$ -adrenergic receptor autoantibodies in refractory hypertension. *PLoS ONE* **3**, e3742 (2008).
36. Noubade, R. *et al.* Activation of p38 MAPK in CD4 T cells controls IL-17 production and autoimmune encephalomyelitis. *Blood* **118**, 3290–3300 (2011).
37. Engel, F. B. *et al.* p38 MAP kinase inhibition enables proliferation of adult mammalian cardiomyocytes. *Genes Dev.* **19**, 1175–1187 (2005).
38. Bettelli, E. *et al.* Myelin oligodendrocyte glycoprotein-specific T cell receptor transgenic mice develop spontaneous autoimmune optic neuritis. *J. Exp. Med.* **197**, 1073–1081 (2003).

Follicular T-helper cell recruitment governed by bystander B cells and ICOS-driven motility

Heping Xu¹, Xuanying Li¹, Dan Liu¹, Jianfu Li^{2,†}, Xu Zhang¹, Xin Chen¹, Shiyue Hou¹, Lixia Peng¹, Chenguang Xu³, Wanli Liu³, Lianfeng Zhang⁴ & Hai Qi¹

Germinal centres support antibody affinity maturation and memory formation¹. Follicular T-helper cells promote proliferation and differentiation of antigen-specific B cells inside the follicle^{2,3}. A genetic deficiency in the inducible co-stimulator (ICOS), a classic CD28 family co-stimulatory molecule highly expressed by follicular T-helper cells, causes profound germinal centre defects^{4,5}, leading to the view that ICOS specifically co-stimulates the follicular T-helper cell differentiation program^{2,6,7}. Here we show that ICOS directly controls follicular recruitment of activated T-helper cells in mice. This effect is independent from ICOS ligand (ICOSL)-mediated co-stimulation provided by antigen-presenting dendritic cells or cognate B cells, and does not rely on Bcl6-mediated programming as an intermediate step. Instead, it requires ICOSL expression by follicular bystander B cells, which do not present cognate antigen to T-helper cells but collectively form an ICOS-engaging field. Dynamic imaging reveals ICOS engagement drives coordinated pseudopod formation and promotes persistent T-cell migration at the border between the T-cell zone and the B-cell follicle *in vivo*. When follicular bystander B cells cannot express ICOSL, otherwise competent T-helper cells fail to develop into follicular T-helper cells normally, and fail to promote optimal germinal centre responses. These results demonstrate a co-stimulation-independent function of ICOS, uncover a key role for bystander B cells in promoting the development of follicular T-helper cells, and reveal unsuspected sophistication in dynamic T-cell positioning *in vivo*.

Follicular T-helper cells are localized in the follicle in part owing to heightened CXCR5 expression³. ICOS-deficient T cells do not upregulate CXCR5 normally and fail to migrate into the follicle (Supplementary Fig. 1). To test whether this localization failure is fully accounted for by inadequate CXCR5 expression, *Icos*^{-/-} or *Icos*^{+/+} OT-II T cells were retrovirally transduced with a CXCR5-expressing vector and activated *in vivo* by immunization with 4-hydroxy-3-nitrophenylacetyl-conjugated ovalbumin (NP-OVA). As shown in Fig. 1a, although many *Icos*^{+/+} OT-II T cells migrated deep into the follicle, few *Icos*^{-/-} T cells were able to do so. CXCR5 overexpression increased the follicular presence of *Icos*^{+/+} OT-II cells. However, CXCR5-transduced *Icos*^{-/-} T cells accumulated towards the border between the T-cell zone and the B-cell follicle (T-B border), but remained scarce deep inside the follicle. This was not due to inadequate CXCR5 complementation, because these *Icos*^{-/-} T cells expressed at least twice as much CXCR5 as their green fluorescent protein (GFP)-transduced wild-type counterpart (Fig. 1b). The ICOS deficiency seems to reduce specifically the efficiency with which T cells re-localize from the T-B border into the follicle (see Supplementary Fig. 2 for typical ICOS expression levels on T cells used in this and other experiments). Quantitatively, we calculated the T-cell density in the T-B bordering region (B_{density}), the density in the adjacent follicle beyond the border (F_{density}), and then deduced the follicular homing coefficient as the ratio between

F_{density} and B_{density} (Fig. 1c). As shown in Fig. 1d, although the follicular homing coefficient of *Icos*^{-/-} T cells was increased by 50% (0.27 versus 0.18, $P < 0.01$) after tenfold CXCR5 overexpression (Fig. 1b), it was still far less than that of *Icos*^{+/+} T cells transduced with GFP (0.69, $P < 0.0001$), in which the CXCR5 levels were 50% lower (Fig. 1b).

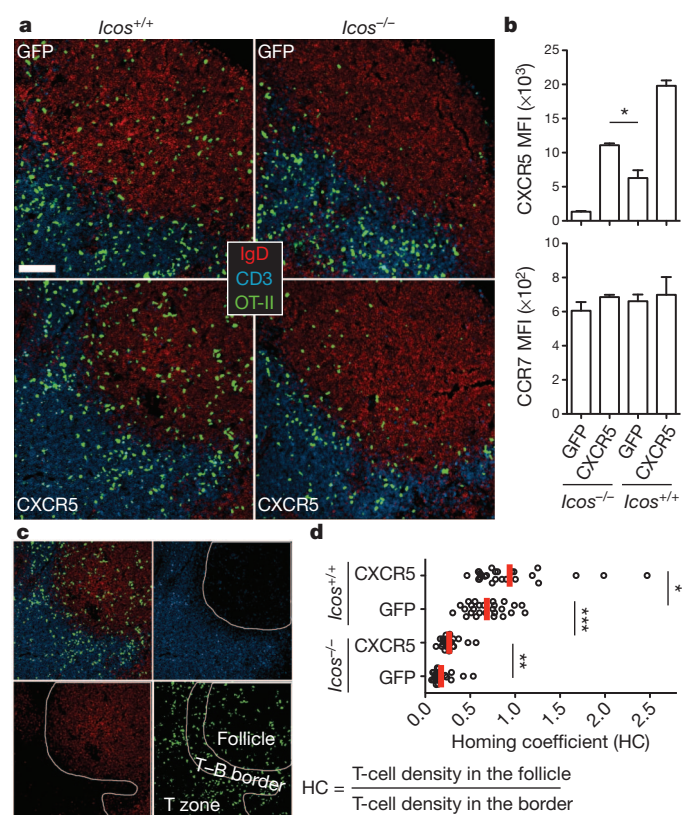


Figure 1 | Follicular recruitment of activated T cells requires ICOS. OT-II T cells of indicated genotypes retrovirally transduced with CXCR5 or control GFP were transferred into B6 mice (5×10^5 per mouse). **a**, OT-II T-cell distribution pattern in lymph nodes 4 days after NP-OVA immunization. Scale bar, 100 μm . **b**, Surface CXCR5 (top) and CCR7 (bottom) expression by flow cytometry (mean \pm s.e.m. of at least three mice per group). MFI, mean fluorescence intensity. **c**, The method to derive the follicular homing coefficient to quantitate homing efficiency. The T-B border region is defined as the region between endogenous CD3⁺ cells farthest into the follicle (white line, top right) and endogenous IgD⁺ cells farthest into the T-cell zone (white line, bottom left). **d**, Homing coefficient (HC) of the four groups of OT-II T cells in **a**. Each symbol represents one follicle and its associated T-B border. Data pooled from three experiments, with at least six lymph nodes from at least three mice per group per experiment. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.0001$.

¹Tsinghua-Peking Center for Life Sciences, Laboratory of Dynamic Immunobiology, School of Medicine, Tsinghua University, Beijing 100084, China. ²Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China. ³School of Life Sciences, Tsinghua University, Beijing 100084, China. ⁴Key Laboratory of Human Disease Comparative Medicine, Ministry of Health, Institute of Laboratory Animal Science, Chinese Academy of Medical Sciences & Comparative Medical Center, Peking Union Medical College, Beijing 100021, China. [†]Present address: Department of Oral and Maxillofacial Surgery, The Methodist Hospital Research Institute, Houston, Texas 77030, USA.

Because CCR7 levels were comparable among all groups (Fig. 1b), there must be an as yet undefined mechanism by which ICOS controls T-cell recruitment from the T–B border into the follicular parenchyma.

To test whether ICOSL co-stimulation provided by priming dendritic cells is involved, adoptively transferred *Icos*^{-/-} or *Icos*^{+/+} OT-II T cells were activated *in vivo* by subcutaneously injected *Icos*^{-/-} or *Icos*^{+/+} dendritic cells that were pre-loaded with the ovalbumin (OVA) immunodominant peptide OVA₃₂₃. As shown in Fig. 2a, whereas *Icos*^{-/-} OT-II T cells were defective in follicular recruitment, wild-type T cells primed by *Icos*^{-/-} dendritic cells were not, indicating that ICOSL co-stimulation by dendritic cells is not responsible for programming the ICOS-dependent follicular recruitment mechanism. Antigen-presenting B cells are another source of co-stimulating ICOSL. To test whether co-stimulation by cognate B cells is involved, OT-II T cells were co-transferred into B6 mice together with *Icos*^{-/-} or *Icos*^{+/+} hen egg lysozyme (HEL)-specific B-cell receptor transgenic MD4 B cells. Four days after immunization with HEL-OVA conjugate antigen, *Icos*^{+/+} but not *Icos*^{-/-} OT-II T cells were abundantly seen in the follicular area, regardless of whether the MD4 B cells could express ICOSL or not (Supplementary Fig. 3). Thus, cognate B cells are not necessarily involved. Bcl6 is a master transcriptional factor that promotes follicular T-helper cell development and thus may drive their follicular localization feature^{2,3}. To test whether Bcl6-mediated programming is an intermediate step in the ICOS-dependent control of follicular T-cell recruitment, Bcl6 was retrovirally transduced into *Icos*^{-/-} or *Icos*^{+/+} OT-II T cells, leading to 40–80-fold overexpression (Supplementary Fig. 4a), and OT-II localization patterns were then assayed after activation *in vivo*. As shown in Fig. 2b, whereas *Icos*^{+/+} OT-II cells were abundantly recruited into the follicle, *Icos*^{-/-} OT-II

cells remained largely blocked from follicular entry despite Bcl6 overexpression. Importantly, Bcl6 overexpression in *Icos*^{+/+} T cells increased the frequency of CXCR5^{hi}PD-1^{hi} follicular T-helper cells and enhanced germinal centre formation (Supplementary Fig. 4b), consistent with previous reports^{8–10}. However, even by the surface phenotyping criteria, Bcl6 overexpression could not correct the follicular T-helper cell defect of *Icos*^{-/-} T cells. Therefore, Bcl6-mediated programming is not an intermediate step in the ICOS-mediated control of follicular T-cell recruitment.

Alternatively, ICOS may function beyond classical co-stimulation and directly promote follicular recruitment of T-helper cells without concomitant antigen receptor signalling-dependent processes. To test this, *Icos*^{+/+} or *Icos*^{-/-} OT-II T cells were activated *in vitro*, transduced with CXCR5, and tested for follicular localization 24 h after transfer into B6 mice that were previously immunized with 4-hydroxy-3-nitrophenylacetyl-conjugated keyhole limpet haemocyanin (NP-KLH) antigen, which cannot stimulate the OT-II T-cell receptor. CXCR5-transduced *Icos*^{+/+} but not *Icos*^{-/-} OT-II cells abundantly migrated deep into the follicle (Fig. 2c), even though the two groups expressed comparable levels of CXCR5 and CCR7 (Supplementary Fig. 5). Similar results were obtained using T cells activated by anti-CD3 and anti-CD28 antibodies before adoptive transfer into naive B6 mice, indicating that neither ICOS-mediated programming *in vitro* nor overt inflammation *in vivo* due to immunization was necessary (Supplementary Fig. 6). Therefore, independent from antigen receptor signalling, ICOS directly promotes follicular recruitment of activated T-helper cells.

To identify the requisite source of ICOSL, we considered follicular bystander B cells, which by definition do not present cognate antigen

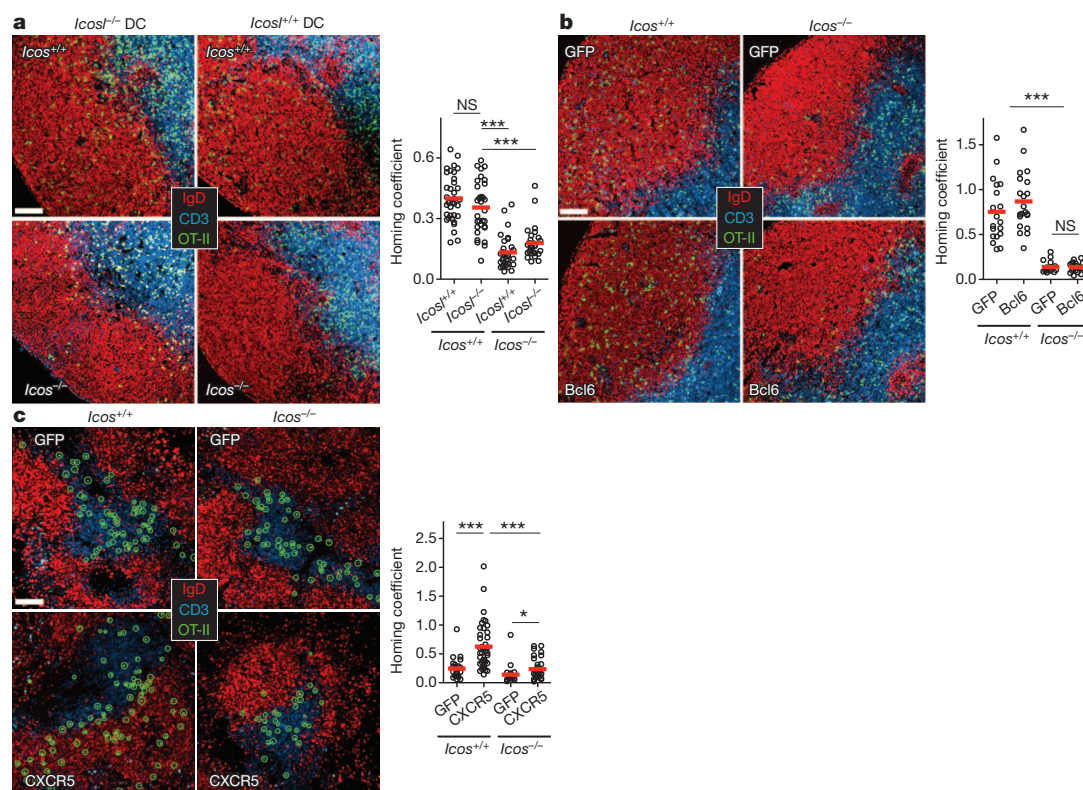


Figure 2 | ICOS-dependent follicular T-cell recruitment does not rely on co-stimulation or Bcl6. **a**, Distribution patterns of *Icos*^{+/+} or *Icos*^{-/-} OT-II T cells in B6 mice, 3–4 days after activation by subcutaneous-injected OVA₃₂₃-pulsed, lipopolysaccharide (LPS)-activated *Icos*^{+/+} or *Icos*^{-/-} dendritic cells (DC). **b**, OT-II T cells were retrovirally transduced with Bcl6 or control GFP, sorted to >90% GFP⁺ purity, and transferred into B6 mice (5 × 10⁵ per mouse). Shown are distribution patterns of OT-II T cells on day 4 after

NP-OVA immunization. **c**, OT-II T cells were transduced with CXCR5 or GFP, and then transferred into B6 mice (3 × 10⁶ sorted GFP⁺ cells per mouse) that had been immunized with the antigen NP-KLH. The OT-II distribution pattern 24 h after transfer is shown (each GFP⁺ cell highlighted with a circle). Scatter plots show homing coefficient analyses as in Fig. 1d, based on data pooled from three (**a**) or two (**b**, **c**) experiments involving 2–3 mice per condition per experiment. Scale bars, 100 μm. *P < 0.05; ***P < 0.0001; NS, not significant.

to T cells, constitutively express ICOSL¹¹, and as an ensemble form an ICOS-engaging field in the follicular parenchyma (Supplementary Fig. 7). To test a potential role for ICOSL on bystander B cells in promoting follicular T-cell recruitment, we constructed 80:20 mixed bone marrow chimaeras using μ MT (B-cell-deficient) and *Icosl*^{-/-} bone marrow cells (see Supplementary Fig. 7 for a diagram) to specifically render B cells ICOSL-deficient in the resulting animals. To control for B-cell-mediated antigen presentation, OVA₃₂₃-pulsed dendritic cells were used to activate the transferred OT-II T cells, and the chimaera between μ MT and *I-A*^{-/-} (also known as *H2-Ab1*^{-/-} or *I-A β* ^{-/-}) B cells was tested in parallel. As shown in Fig. 3, whereas activated OT-II T cells were abundantly recruited into follicles composed of wild-type or *I-A*^{-/-} B cells, they were essentially absent in ICOSL-deficient follicles. These results suggest that, by constitutively expressing ICOSL without co-displaying agonistic peptide-major histocompatibility complex (MHC) complexes, follicular bystander B cells engage ICOS on activated T-helper cells that reach the T-B border, and promote their subsequent follicular recruitment in a co-stimulation-independent manner.

How ICOS triggering by the ensemble of bystander B cells promotes the directional outcome of follicular T-cell recruitment from the T-B border remains unclear. Whereas follicular re-localization depends on the net vector of CXCR5- and CCR7-mediated direction-sensing¹², ICOS exerts its effects independently of CXCR5 and CCR7 expression levels (Figs 1 and 2c) and without changing their sensitivities to respective chemokine ligands (Supplementary Fig. 8). Conversely, directional migration depends not only on gradient sensing but also on persistent random motility. This latter process is characterized by coordinated pseudopod generation in a polarized manner, either spontaneously or in response to a uniform field, drives cells to move persistently for a certain period before random direction change and repetition of the motility^{13,14}. As a result, a gradient steers cell migration by imposing a directional bias onto persistent random motility, without which cells may sense but cannot efficiently move towards the desired direction^{15,16}. We thus sought to test whether ICOS triggering promotes persistent motility of T cells and thereby facilitates follicular recruitment from the T-B border.

Similar to what was described for human T cells^{17,18}, anti-ICOS antibody treatment led to mouse T-cell polarization (Supplementary

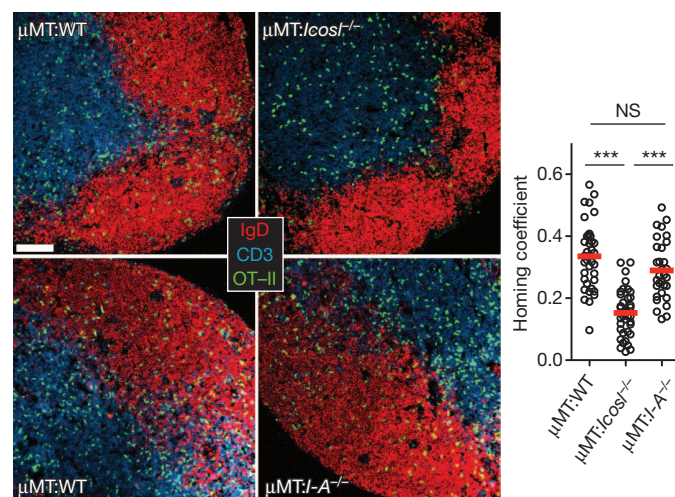


Figure 3 | ICOSL expressed on bystander B cells is required for follicular T-helper cell recruitment. Mixed bone marrow chimaeras of indicated types received 10^6 naive GFP-expressing OT-II T cells per mouse, and were then subcutaneously injected with 2×10^6 OVA₃₂₃-pulsed dendritic cells. Representative OT-II tissue distribution patterns 4 days later and quantitative homing coefficient analyses as in Fig. 1d are shown. Data represent three sets of μ MT:*Icosl*^{-/-} chimaeras and two sets of μ MT:*I-A*^{-/-} chimaeras. WT, wild type. Scale bar, 100 μ m. ****P* < 0.0001.

Fig. 9). To analyse motility *in vitro*, T cells were transduced with a Lifeact-based F-actin-binding reporter¹⁹ and examined by total internal reflection fluorescence (TIRF) microscopy on lipid bilayers that can or cannot trigger ICOS. Under the control condition, T cells exhibit spontaneous actin dynamics, periodically in a wave-like form along the membrane to extend pseudopodial extensions in random directions, whereas the cell body appeared round and without overt displacement (Fig. 4a and Supplementary Videos 1 and 2). In marked contrast, on the ICOS-triggering bilayer, T cells exhibited leading edges, from which actin waves drove pseudopods in a coordinated left-right fashion, enabling T cells to move persistently (Fig. 4a and Supplementary Videos 3 and 4). This was reminiscent of spontaneous motility described for amoebic *Dictyostelium* and of that induced in neutrophils by a uniform field¹⁴. Quantitatively, ICOS triggering markedly increased cell displacement (Fig. 4b), centroid speed (by ~40%; Fig. 4c, left), and directional persistence (by ~200%; Fig. 4c, right). This motility-driving effect required activities of phosphatidylinositol-3-OH kinase (PI(3)K), as it was blocked by treatment with p110 δ -specific PI(3)K inhibitor CAL-101 (Supplementary Fig. 10 and Supplementary Videos 5 and 6). Notably, CD28 triggering, which could also activate the PI(3)K pathway albeit to a lesser extent²⁰, failed to drive persistent motility even though it increased T-cell membrane dynamics and polarization (Supplementary Fig. 10 and Supplementary Videos 6 and 7). Therefore, ICOS triggering uniquely enhances coordinated pseudopod dynamics in T cells and promotes their persistent motility in a PI(3)K-dependent manner.

To test whether ICOS triggering promotes persistent motility *in vivo*, particularly as governed by bystander B cells at the T-B border, we conducted two-photon intravital imaging analysis (Supplementary Fig. 11). To capture T cell pseudopod dynamics in detail, we conducted imaging at one frame per 10 s, with the x-y pixel size at $0.49 \times 0.49 \mu$ m. Activated OT-II T cells at the T-B border were seen to extend pseudopods from the leading edge in a left-right coordinated fashion, approximately every 10–40 s (Supplementary Fig. 12 and Supplementary Video 8). Only very occasionally (4–5% of time) would they pause, exhibit a shape index of less than 2, and display no pseudopod-like protrusions, that is, in a depolarized state (see also Methods and Supplementary Fig. 11). When *Icosl*^{-/-} and wild-type OT-II T cells reaching the same T-B border were compared, the former was much more likely to exhibit the depolarized state and to remain so for a longer period time (Fig. 4d, e and Supplementary Videos 9–11). Consistent with this decrease in pseudopod dynamics at the single-cell level, *Icosl*^{-/-} T cells as a population exhibited reductions in displacement kinetics, speed and persistence (Fig. 4f–h and Supplementary Videos 9–11; see Supplementary Fig. 13 for details on statistical analyses). These data suggest ICOS is required for persistent motility of T cells at the T-B border. To determine whether ICOSL on bystander B cells is indeed the driving force, we conducted imaging with μ MT:*Icosl*^{-/-} (80:20)-mixed bone marrow chimaeras as recipients. In these hosts, the frequency of wild-type OT-II T cells in the depolarized state increased from ~5% to ~19%, indistinguishable from their *Icosl*^{-/-} counterparts reaching the same T-B border (Fig. 4i, j and Supplementary Video 12). As a result, the two cell populations became comparable in displacement kinetics and persistence, with *Icosl*^{-/-} T cells now being slightly faster (Fig. 4k–m and Supplementary Video 12). Collectively, these data suggest that ICOS triggering *in vivo* by the ensemble of follicular bystander B cells enhances pseudopod dynamics of activated T cells, increases their persistent motility, and can thereby promote T-cell recruitment from the T-B border into the follicle.

Finally, to verify the functional significance of this ICOS-dependent, bystander B-cell-mediated T-cell recruitment mechanism for the germinal centre response and follicular T-helper cell development, μ MT:*Icosl*^{+/+} and μ MT:*Icosl*^{-/-} (80:20)-mixed bone marrow chimaeras were compared for competency to host an adoptive germinal centre response by MD4 B cells collaborating with OT-II T cells after

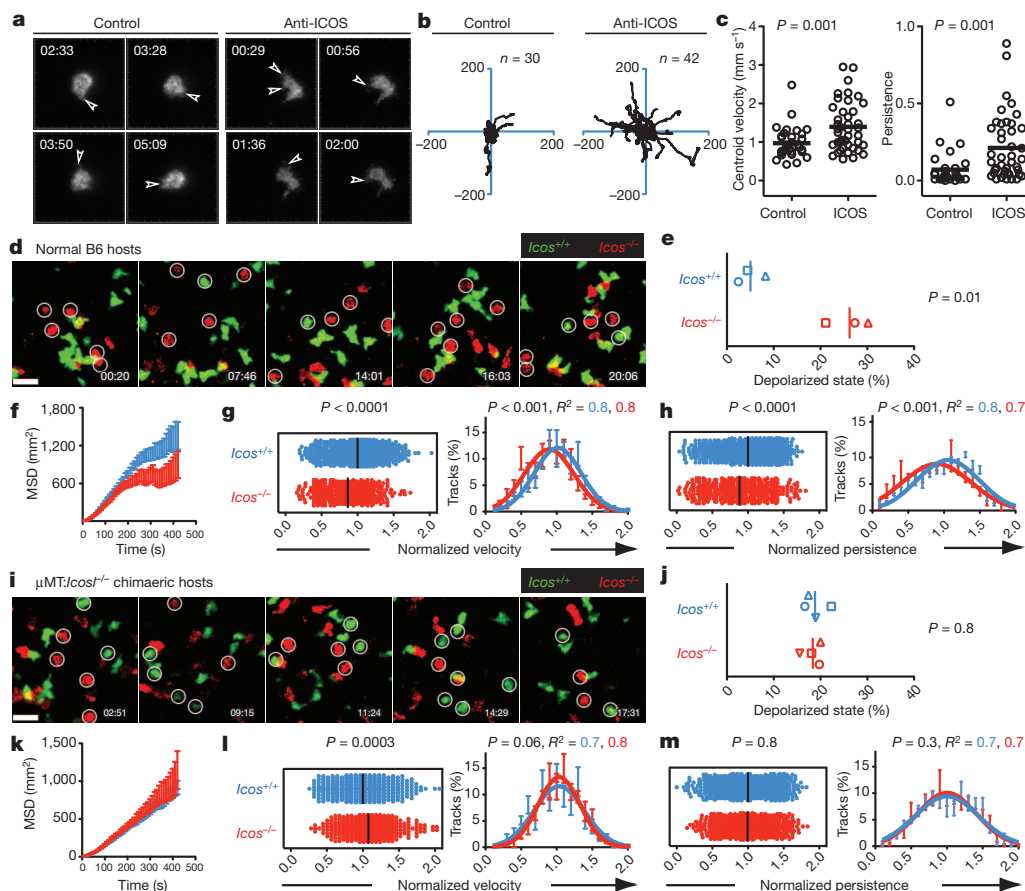


Figure 4 | ICOS-driven persistent T cell motility *in vitro* and *in vivo*.

a–c, Lifeact-labelled T cells were imaged by TIRF microscopy on lipid bilayers that display isotype control or anti-ICOS antibody. Cells were tracked for 8–10 min at one frame per second. **a,** Representative images of T cells (see also Supplementary Videos 1–4). Arrowheads indicate actin waves and pseudopod extensions. **b,** x – y displacement (in μm) plots of individual cell traces, with starting positions re-aligned at the same origin. **c,** The T-cell centroid velocity (left) and directional persistence (right) as measured by path length-normalized displacement. Lines denote the mean. **d–m,** GFP-expressing $Icos^{+/+}$ and dsRed-expressing $Icos^{-/-}$ OT-II T cells were visualized by two-photon intravital microscopy at the T–B border in normal B6 (**d–h**; 601 $Icos^{+/+}$ and 344 $Icos^{-/-}$ tracks from three experiments) or $\mu\text{MT}:Icos^{-/-}$ chimaeric (**i–m**; 538 $Icos^{+/+}$ and 466 $Icos^{-/-}$ tracks from four experiments) hosts 3 days after activation by OVA₃₂₃-pulsed dendritic cells. All tracks are of duration

between 2 and 7 min. **d, i,** Time-lapse images showing a typical range of cell morphology. Circles highlight the depolarized state. Scale bars, 20 μm .

e, j, Frequencies of depolarized states exhibited by $Icos^{+/+}$ and $Icos^{-/-}$ T cells (>100 cells counted for each experiment represented by each symbol, and symbols of the same shape represent matched measurements; lines denote the means). **f, k,** Mean squared displacement (MSD) over a period of 7 min. Data are mean \pm s.d. **g, l,** Scatter plots of velocities pooled from all experiments (left, lines denote the means, P values are from Student's t -test), and Gaussian fits for velocity distributions of individual experiments (right, P values from the extra sum-of-squares F -test, R^2 for goodness of fit, error bars denote histogram s.e.m. of 3–4 experiments). **h, m,** Scatter plots and Gaussian fits for cell persistence measured as the path-length-normalized displacement. Statistical analyses and presentation are as in **g** and **l**.

HEL-OVA immunization. As shown in Supplementary Fig. 14, intrinsically competent OT-II T cells failed to promote wild-type MD4 B cells to generate a normal germinal centre response, and failed to produce a normal frequency of CXCR5^{hi}PD-1^{hi} follicular T-helper cells in the $\mu\text{MT}:Icos^{-/-}$ host. Conversely, wild type:wild type or wild type: $Icos^{-/-}$ chimaeras were comparable hosts for the germinal centre and follicular T-helper cell response (Supplementary Fig. 15), ruling out an effect of $\sim 20\%$ general reduction in ICOSL-expressing cells. Very specifically, therefore, ICOSL-expressing follicular bystander B cells are essential for promoting optimal collaboration between antigen-specific T and B cells that gives rise to normal follicular T-helper cell and germinal centre development.

This study establishes a co-stimulation-independent function for ICOS in regulating follicular T-helper cell recruitment. This function requires PI(3)K activities, and is probably based on feeding the membrane phosphatidylinositol-3,4,5-triphosphate cycle that triggers actin waves and polarized pseudopod formation^{14,15,21}. The requirement of ICOS for optimal T-cell motility *in vivo* is specific to the T–B border region (presumably also inside the follicle), as $Icos^{-/-}$ T cells have no defect in migrating in the T-cell zone (Supplementary Fig. 16 and

Supplementary Videos 13 and 14; see Supplementary Note 1 for further discussion on motility). By promoting persistent motility that a follicle-directed chemosensing bias would rely on to drive efficient homing, ICOS serves as a 'license' for T cells to take the follicular residence. Human resting B cells express and further upregulate ICOSL after inflammatory cytokine stimulation²², and the co-stimulation-independent ICOS function is probably a shared feature between mouse and human. Given the wide non-lymphoid expression of ICOSL²³, ICOS-driven T-cell motility might have a role in orchestrating peripheral inflammation²⁴.

This study also establishes a key role for bystander follicular B cells in facilitating the follicular T-helper cell and germinal centre development. By forming the tightly packed follicular parenchyma, these cells present an ICOS-engaging field to incoming T-helper cells, promoting their initial recruitment from the T–B border for eventual development into follicular T-helper cells inside the follicle. This is in marked parallel to the previous finding that SAP (SLAM-associated protein)-dependent cognate T–B cell interactions are essential for germinal centre recruitment and retention of follicular T-helper cells²⁵, thus re-emphasizing the B-cell compartment as the driver for the

development of their own helpers. Follicular B cells hence have a dual role in the T-cell-dependent B-cell response: they serve not only as a clonal repertoire of individualistic B cells in search of their own antigen, but also as an altruistic ensemble that facilitates the delivery of T-cell help to antigen-engaged fellow B cells.

METHODS SUMMARY

Essential methods were previously described^{25,26}. For quantitative analysis of polarized and depolarized states of T cells migrating at the T–B border, six time frames equally spaced in temporal order were taken from each 20-min image sequence. At each time frame, only cells that exhibit a shape index²⁷ (cell length divided by its width at the middle along the length) of no less than 2—or display at least one pseudopod protrusion when shape index is less than 2—are considered polarized (see Supplementary Fig. 11 for examples).

Full Methods and any associated references are available in the online version of the paper.

Received 8 November 2012; accepted 5 March 2013.

- Victora, G. D. & Nussenzweig, M. C. Germinal centers. *Annu. Rev. Immunol.* **30**, 429–457 (2012).
- Crotty, S. Follicular helper CD4 T cells (T_{FH}). *Annu. Rev. Immunol.* **29**, 621–663 (2011).
- Vinuesa, C. G. & Cyster, J. G. How T cells earn the follicular rite of passage. *Immunity* **35**, 671–680 (2011).
- Dong, C., Temann, U. A. & Flavell, R. A. Cutting edge: critical role of inducible costimulator in germinal center reactions. *J. Immunol.* **166**, 3659–3662 (2001).
- McAdam, A. J. *et al.* ICOS is critical for CD40-mediated antibody class switching. *Nature* **409**, 102–105 (2001).
- Gigoux, M. *et al.* Inducible costimulator promotes helper T-cell differentiation through phosphoinositide 3-kinase. *Proc. Natl Acad. Sci. USA* **106**, 20371–20376 (2009).
- Choi, Y. S. *et al.* ICOS receptor instructs T follicular helper cell versus effector cell differentiation via induction of the transcriptional repressor Bcl6. *Immunity* **34**, 932–946 (2011).
- Johnston, R. J. *et al.* Bcl6 and Blimp-1 are reciprocal and antagonistic regulators of T follicular helper cell differentiation. *Science* **325**, 1006–1010 (2009).
- Nurieva, R. I. *et al.* Bcl6 mediates the development of T follicular helper cells. *Science* **325**, 1001–1005 (2009).
- Yu, D. *et al.* The transcriptional repressor Bcl-6 directs T follicular helper cell lineage commitment. *Immunity* **31**, 457–468 (2009).
- Yoshinaga, S. K. *et al.* T-cell co-stimulation through B7RP-1 and ICOS. *Nature* **402**, 827–832 (1999).
- Haynes, N. M. *et al.* Role of CXCR5 and CCR7 in follicular Th cell positioning and appearance of a programmed cell death gene-1^{high} germinal center-associated subpopulation. *J. Immunol.* **179**, 5099–5108 (2007).
- Van Haastert, P. J. & Devreotes, P. N. Chemotaxis: signalling the way forward. *Nature Rev. Mol. Cell Biol.* **5**, 626–634 (2004).
- Iglesias, P. A. & Devreotes, P. N. Biased excitable networks: how cells direct motion in response to gradients. *Curr. Opin. Cell Biol.* **24**, 245–253 (2012).
- Kay, R. R., Langridge, P., Traynor, D. & Hoeller, O. Changing directions in the study of chemotaxis. *Nature Rev. Mol. Cell Biol.* **9**, 455–463 (2008).
- Insall, R. H. Understanding eukaryotic chemotaxis: a pseudopod-centred view. *Nature Rev. Mol. Cell Biol.* **11**, 453–458 (2010).
- Okamoto, N. *et al.* AILIM/ICOS signaling induces T-cell migration/polarization of memory/effector T-cells. *Int. Immunol.* **16**, 1515–1522 (2004).
- Nukada, Y. *et al.* AILIM/ICOS-mediated elongation of activated T cells is regulated by both the PI3-kinase/Akt and Rho family cascade. *Int. Immunol.* **18**, 1815–1824 (2006).
- Riedl, J. *et al.* Lifeact: a versatile marker to visualize F-actin. *Nature Methods* **5**, 605–607 (2008).
- Parry, R. V., Rumbley, C. A., Vandenberghe, L. H., June, C. H. & Riley, J. L. CD28 and inducible costimulatory protein Src homology 2 binding domains show distinct regulation of phosphatidylinositol 3-kinase, Bcl-x_L, and IL-2 expression in primary human CD4 T lymphocytes. *J. Immunol.* **171**, 166–174 (2003).
- Arai, Y. *et al.* Self-organization of the phosphatidylinositol lipids signaling system for random cell migration. *Proc. Natl Acad. Sci. USA* **107**, 12399–12404 (2010).
- Yoshinaga, S. K. *et al.* Characterization of a new human B7-related protein: B7RP-1 is the ligand to the co-stimulatory protein ICOS. *Int. Immunol.* **12**, 1439–1447 (2000).
- Swallow, M. M., Wallin, J. J. & Sha, W. C. B7h, a novel costimulatory homolog of B7.1 and B7.2, is induced by TNF α . *Immunity* **11**, 423–432 (1999).
- Simpson, T. R., Quezada, S. A. & Allison, J. P. Regulation of CD4 T cell activation and effector function by inducible costimulator (ICOS). *Curr. Opin. Immunol.* **22**, 326–332 (2010).
- Qi, H., Cannons, J. L., Klauschen, F., Schwartzberg, P. L. & Germain, R. N. SAP-controlled T–B cell interactions underlie germinal centre formation. *Nature* **455**, 764–769 (2008).
- Liu, W., Won Sohn, H., Tolar, P., Meckel, T. & Pierce, S. K. Antigen-induced oligomerization of the B cell receptor is an early target of Fc γ RIIB inhibition. *J. Immunol.* **184**, 1977–1989 (2010).
- Miller, M. J., Wei, S. H., Parker, I. & Cahalan, M. D. Two-photon imaging of lymphocyte motility and antigen response in intact lymph node. *Science* **296**, 1869–1873 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements H.Q. is indebted to his late mother J. Wang for inspiration, and thanks Y. Hong for support. H.Q. was a Tsinghua-Bayer Investigator and a Tsinghua-Janssen Investigator. This work was funded in part by the National Natural Science Foundation of China (grant no. 81072464 and 81161120405), the Program for New Century Excellent Talents in University (no. 20091042000), Tsinghua-Yu-Yuan Medical Sciences Fund (no. 20240000585), Tsinghua University Initiative Scientific Research Program (no. 2010Z02150), Institut Mérieux, and the Tsinghua-Peking Center for Life Sciences.

Author Contributions H.X. conducted most of the experiments. H.Q. conceptualized the study, designed the experiments with input from all authors, and wrote the paper. All authors contributed collectively to interpreting data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.Q. (qihai@biomed.tsinghua.edu.cn).

METHODS

Essential technical information was previously described^{25,26}. For quantitative analysis of polarized and depolarized states of T cells migrating at the T–B border, six time frames equally spaced in temporal order were taken from each 20-min image sequence. At each time frame, only cells that exhibit a shape index²⁷ (cell length divided by its width at the middle along the length) of no less than 2—or display at least one pseudopod protrusion when shape index is less than 2—are considered polarized (see Supplementary Fig. 11 for examples).

Mice. B6 (Jax 664), μ MT (Jax 2288), $Icos^{-/-}$ (Jax 4859), $Icosl^{-/-}$ (Jax 4657), GFP-expressing (Jax 4353), dsRed-expressing (Jax 6051), OVA_{323–339}-specific T-cell receptor transgenic OT-II (Jax 4194), and HEL-specific Ig-transgenic MD4 (Jax 2595) mice were from the Jackson Laboratory. $I-A^{-/-}$ (also known as $H2-Ab1^{-/-}$ or $I-A\beta^{-/-}$) mice were from Taconic Farms. Relevant mice were interbred to obtain GFP-OT-II, $Icos^{+/+}$ or $Icos^{-/-}$ dsRed-OT-II, and $Icosl^{+/+}$ or $Icosl^{-/-}$ dsRed-MD4 mice. All mice were maintained under specific pathogen-free conditions, and used in accordance of governmental and institutional guidelines for animal welfare.

Cell culture, retrovirus and *in vitro* transduction. Naive OT-II T cells, MD4 B cells, or polyclonal B cells were isolated using the negative CD4 T cell isolation kit, or the naive B cell isolation kit (Miltenyi Biotec), according to the manufacturer's protocols. Dendritic cells were isolated by CD11c microbeads (Miltenyi Biotec) from mouse spleens after digestion with 400 μ g ml⁻¹ liberase CI and 20 μ g ml⁻¹ DNase I for 30 min (Roche). Retroviruses expressing desired target genes were packaged with the Plate-E system. The MSCV-based, GFP-tagged vector used throughout this study was custom-modified from the pMSCVpuro vector (Clontech) by substituting its phosphoglycerate kinase promoter-puromycin cassette with a human ubiquitin promoter-driven enhanced GFP fragment. For *in vitro* T-cell activation, OT-II T cells were co-cultivated with mitomycin-treated splenocytes in the presence of 2 μ M OVA_{323–339} peptide (Genscript). For retroviral transduction, 3×10^5 activated OT-II cells were spin-infected at 1,500g with appropriate viral supernatants in the presence of 1 μ g ml⁻¹ polybrene (Sigma) and 10 ng ml⁻¹ IL-2 (Peprotech) for 2 h at 32 °C. Infected T cells were then transferred into new wells with fresh media supplemented with 10 ng ml⁻¹ IL-2 for further culture with splits as needed.

Adoptive transfer, antigen and immunization. To activate OT-II T cells in draining lymph nodes, naive or *in vitro* activated and retrovirally transduced OT-II T cells were intravenously transferred into B6 or mixed chimaeric hosts before subcutaneous immunization with 30 μ g OVA protein (Sigma) or NP-OVA (Biosearch Tech) plus 0.5 μ g LPS in alum (Thermo Scientific). For certain experiments, protein immunization was replaced with subcutaneous injection of dendritic cells that were pulsed with 3 μ M OVA₃₂₃ peptide in the presence of 0.2 μ g ml⁻¹ LPS for 2 h at 37 °C and then repeatedly washed. When transduced T cells were used, they were parked in the adoptive host for 4 days before re-activation *in vivo*. For inducing endogenous immune responses in the spleen, intraperitoneal injection of 100 μ g NP-KLH (Biosearch Tech) mixed with 1 μ g LPS (Sigma) in alum was used. To induce germinal centre formation by MD4 B cells with help from OT-II T cells, subcutaneous injection of 30 μ g HEL-OVA conjugate antigen was used. The HEL-OVA conjugate was prepared using a HydraLink conjugation kit (SoluLink) as previously described²⁵.

Construction of bone marrow chimaeras. B6 recipients were lethally irradiated by X-ray (5 Gy \times 2), and then intravenously transferred with a combination of 4×10^6 bone-marrow leukocytes from indicated donors mixed according to indicated ratios. Chimaeras were used for experiments 8 weeks after the initial reconstitution.

Flow cytometry and immunohistochemistry. To phenotype OT-II T cells and/or MD4 B cells in the lymph node by flow cytometry, lymph node cells were washed, incubated with a mixture of 10% goat and rabbit serum and 20 μ g ml⁻¹ 2.4G2 (BioXcell), and then stained with indicated monoclonal antibodies in MACS buffer (PBS supplemented with 1% FBS and 5 mM EDTA). Staining reagents included Alexa Fluor 700 anti-CD4, allophycocyanin (APC)-Cy7-anti-CD19, biotinylated anti-ICOS, biotinylated anti-ICOSL, phycoerythrin (PE)-anti-ICOSL, biotinylated anti-CCR7, streptavidin-APC, streptavidin-PE, and DyLight649 goat anti-hamster IgG antibody purchased from Biolegend; FITC anti-GL7 and purified anti-PD-1 from eBioscience; PE-cy7-anti-CD95, PE-anti-I-A, APC-anti-CD11c, Alexa Fluor 647 anti-Bcl6, biotinylated anti-IgM α , PE-anti-ICOS, and biotinylated anti-CXCR5 from BD Biosciences. Isotype-matched nonspecific antibodies were also purchased from these companies. Cells were stained on ice with primary reagents for 60–90 min followed by staining with secondary reagents for 30 min. Data were collected on a LSR II cytometer (BD Biosciences) and analysed with FlowJo software (TreeStar). Dead cells and non-singlet events were excluded from analyses based on 7-amino-actinomycin D (7-AAD) staining (Biotium) and the forward scatter area (FSC-A) versus height (FSC-H) characteristics. To

examine T-cell distribution patterns *in vivo*, immunohistochemical staining of lymph node sections was conducted according to protocols previously described²⁸. When the follicular homing coefficient was quantified as detailed in Fig. 1d, non-consecutive sections were used to avoid the same follicle and its associated T–B border being repeatedly counted. Staining reagents included eFluor450 anti-CD3 (eBioscience), Alexa Fluor 647 anti-B220 (BD), Alexa Fluor 647 anti-IgD (eBioscience), purified anti-ICOSL and Alexa Fluor 568 anti-rat IgG (Invitrogen). Slides were mounted with the ProlongGold Antifade reagent (Invitrogen) and examined with an Olympus FV1000 upright microscope using $\times 20$ air or $\times 40$ oil immersion lens.

Transwell migration assay. T cells of indicated genotypes were rested in the RPMI medium containing 1% FBS at 37 °C for 1 h before being loaded as a $100\text{-}\mu$ l suspension of 10^5 cells into the upper transwell in a 96-well-plate format (5- μ m pore, Corning). The cell suspension was supplemented with the anti-ICOS or isotype-matched control antibody at a final concentration of 10 μ g ml⁻¹. Recombinant CXCL13 or CCL21 (Peprotech) of indicated concentrations was added to the bottom wells before the cells were allowed to trans-migrate for 3 h at 37 °C in an incubator. Cells that have migrated to the bottom wells were enumerated by flow cytometry with an internal counting standard.

Imaging of T-cell actin cytoskeleton and motility *in vitro*. To examine actin polymerization and the polarization state of T cells, OT-II T-cell blasts were incubated with 5 μ g ml⁻¹ control or anti-ICOS antibody (Biolegend) at 4 °C for 1 h, washed twice with PBS containing 1% serum, dropped onto poly-L-lysine-coated glass (Electron Microscopy Science) for incubation at 37 °C for 10 min, fixed with 1% paraformaldehyde and permeabilized with 0.1% saponin, stained for F-actin with 10 μ M Alexa Fluor 647-conjugated phalloidin, mounted with the ProlongGold Antifade reagent (Invitrogen), and then imaged with a Nikon Ti-E microscope equipped with $\times 100$ oil immersion lens. To examine cytoskeleton dynamics and motility of T cells, TIRF microscopy was conducted using a biotinylated lipid bilayer system as previously described with minor modifications²⁶. In brief, planar bilayers were made by mixing 99% 1,2-dioleoyl-*sn*-glycero-3-phosphocholine lipid and 1% 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine-cap-biotin (Avanti Polar Lipids). After sonication and ultracentrifugation of the mixed lipids, a 0.1 mM solution of the formed unilamellar vesicles was placed on pre-cleaned glass coverslips for 15 min at room temperature. The coverslip was then sequentially incubated with 50 nM streptavidin (Jackson ImmunoResearch) for 15 min and 5 μ g ml⁻¹ control or anti-ICOS or anti-CD28 antibody (Biolegend) for 15 min with PBS washings in-between. After blocking with 1% serum at 37 °C for 30 min, the bilayers were used for TIRF imaging. T cells transduced with the Lifeact-mRuby vector were suspended in 1% FBS-supplemented RPMI media and then placed onto the bilayer and imaged in custom-made physiology chamber at 37 °C. For PI(3)K inhibition, the p110 δ -selective inhibitor CAL-101 (ref. 29; Selleck Chemicals) was used at 1 μ M to pre-treat T cells overnight and maintained in the imaging buffer. TIRF images were acquired at one frame per second using an Olympus IX-81 microscope equipped with a TIRF port, a 512 \times 512 EMCCD camera (Andor), a 561-nm laser (Coherent), and $\times 100$ 1.45 numerical aperture (NA) objective lens (Olympus). Acquisition was controlled by Metamorph software (MDS Analytical Technologies), and data were processed by NIH ImageJ and Imaris (Bitplane).

Two-photon intravital imaging of T-cell motility *in vivo*. Basic procedures for intravital lymph node imaging were essentially as previously described²⁵. The imaging system was composed of a MaiTai DeepSee laser (Spectra-Physics) and an Olympus FV1000 upright microscope equipped with the XLPlan $\times 25$ water immersion lens (NA 1.05, Olympus). The motorized stage on which live mice were imaged was enclosed in a customized chamber that was heated to 37 °C at equilibrium. To visualize T cells of different *Icos* genotypes in the same region after activation *in vivo*, OT-II T cells that carry germline-transgene GFP and dsRed were used. This was the only workable combination of fluorescent proteins for our experiments, because alternatives would have to involve the cyan fluorescent protein (CFP) transgenic line that, albeit successfully used for germinal centre imaging before²⁵, was not sufficiently 'bright' for detailed analysis of T cells in the generally deeper region of the T–B border. To locate follicles, wild-type or, in the case of μ MT:*Icosl*^{-/-} recipient hosts, *Icosl*^{-/-} naive B cells labelled with 50–100 μ M CMF₂HC (Invitrogen) were intravenously transferred 24 h before imaging. The combination of CMF₂HC-labelled B cells and GFP T cells imaged at 800 nm allowed identification of the T–B border, and care was taken to choose fields in which the border segregation was evident on the *x*–*y* plane (also see Supplementary Fig. 9 for details). Actual experiments for imaging GFP and dsRed were conducted at 880–920 nm. To capture pseudopod dynamics of migrating T cells at the T–B border, imaging was conducted with a zoom factor of 2 to achieve an *x*–*y* pixel size of ~ 0.5 μ m and at a time resolution of 10 s for each frame, yielding a typical spatial dimension of the imaged volume at 250 \times 250 \times 21 μ m. Each image sequence was 20 min in duration. After acquisition, four-dimensional

data sets were analysed using Imaris software (Bitplane). Cell migration was analysed by automatic Imaris cell tracking module aided with manual supervision and verification. Cell tracks that lasted for less than 2 min were excluded from analysis. Analysis of mean squared displacements only included data from the first 7 min, because less than 10% of tracks lasted longer than 7 min within the imaged volume, making mean displacement data beyond 7 min less representative. Persistence is measured as displacement-normalized path length. When T-zone motility was analysed, naive or T cells activated *in vitro* were labelled with 50 μ M CMF₂HC or 2 μ M CMFDA (Invitrogen). Adobe Photoshop, AfterEffect and Illustrator were used to prepare presentations of time-lapse image sequences

and videos, which are played back at 20 frames per second unless indicated otherwise.

Statistical analysis. Unless specifically indicated otherwise, *t*-tests were used to compare endpoint means of different groups. Statistical tests, nonlinear regression, and graphing were done with Prism (GraphPad).

28. Qi, H., Egen, J. G., Huang, A. Y. & Germain, R. N. Extrafollicular activation of lymph node B cells by antigen-bearing dendritic cells. *Science* **312**, 1672–1676 (2006).
29. Norman, P. Selective PI3K δ inhibitors, a review of the patent literature. *Expert Opin. Ther. Pat.* **21**, 1773–1790 (2011).

High-level semi-synthetic production of the potent antimalarial artemisinin

C. J. Paddon¹, P. J. Westfall^{1†}, D. J. Pitera¹, K. Benjamin¹, K. Fisher¹, D. McPhee¹, M. D. Leavell¹, A. Tai¹, A. Main^{1†}, D. Eng¹, D. R. Polichuk², K. H. Teoh^{2†}, D. W. Reed², T. Treynor¹, J. Lenihan^{1†}, M. Fleck¹, S. Bajad^{1†}, G. Dang^{1†}, D. Dengrove¹, D. Diola¹, G. Dorin¹, K. W. Ellens^{2†}, S. Fickes¹, J. Galazzo¹, S. P. Gaucher¹, T. Geistlinger¹, R. Henry¹, M. Hepp^{2†}, T. Horning¹, T. Iqbal¹, H. Jiang¹, L. Kizer¹, B. Lieu¹, D. Melis¹, N. Moss¹, R. Regentin^{1†}, S. Secrest¹, H. Tsuruta¹, R. Vazquez¹, L. F. Westblade¹, L. Xu¹, M. Yu¹, Y. Zhang^{2†}, L. Zhao¹, J. Lievens^{1†}, P. S. Covelio², J. D. Keasling^{3,4,5,6}, K. K. Reiling¹, N. S. Renninger¹ & J. D. Newman¹

In 2010 there were more than 200 million cases of malaria, and at least 655,000 deaths¹. The World Health Organization has recommended artemisinin-based combination therapies (ACTs) for the treatment of uncomplicated malaria caused by the parasite *Plasmodium falciparum*. Artemisinin is a sesquiterpene endoperoxide with potent antimalarial properties, produced by the plant *Artemisia annua*. However, the supply of plant-derived artemisinin is unstable, resulting in shortages and price fluctuations, complicating production planning by ACT manufacturers². A stable source of affordable artemisinin is required. Here we use synthetic biology to develop strains of *Saccharomyces cerevisiae* (baker's yeast) for high-yielding biological production of artemisinic acid, a precursor of artemisinin. Previous attempts to produce commercially relevant concentrations of artemisinic acid were unsuccessful, allowing production of only 1.6 grams per litre of artemisinic acid³. Here we demonstrate the complete biosynthetic pathway, including the discovery of a plant dehydrogenase and a second cytochrome that provide an efficient biosynthetic route to artemisinic acid, with fermentation titres of 25 grams per litre of artemisinic acid. Furthermore, we have developed a practical, efficient and scalable chemical process for the conversion of artemisinic acid to artemisinin using a chemical source of singlet oxygen, thus avoiding the need for specialized photochemical equipment. The strains and processes described here form the basis of a viable industrial process for the production of semi-synthetic artemisinin to stabilize the supply of artemisinin for derivatization into active pharmaceutical ingredients (for example, artesunate) for incorporation into ACTs. Because all intellectual property rights have been provided free of charge, this technology has the potential to increase provision of first-line antimalarial treatments to the developing world at a reduced average annual price.

Before the discovery of the enzymes that complete the biosynthetic pathway of artemisinin production (see Supplementary Fig. 1 for a complete overview), several improvements were made to the original amorphadiene-producing strain Y337 (ref. 3). We replaced the *MET3* promoter with the copper-regulated *CTR3* promoter (Fig. 1a), enabling restriction of *ERG9* expression (*ERG9* encodes squalene synthase, which catalyses the competing reaction of joining two farnesyl diphosphate moieties to form squalene) by addition of the inexpensive repressor CuSO_4 to the medium rather than the more expensive methionine^{4–6}. Strains Y1516 (*P_{CTR3}-ERG9*) and Y337 (*P_{MET3}-ERG9*)

(Supplementary Table 1) both produced similar amounts of amorphadiene (Supplementary Fig. 2), demonstrating the equivalence of the *MET3* and *CTR3* promoters for repression of *ERG9* expression. We compared the production of amorphadiene from Y337 with the production of artemisinic acid from Y285, a variant of Y337 that also expressed the amorphadiene oxidase *CYP71AV1* (a cytochrome P450) and *A. annua CPR1* (its cognate reductase) from a high-copy plasmid (pAM322)³. Both strains were grown in a fed-batch fermentor with mixed glucose and ethanol feed. Whereas Y337 produced more than 12 g l^{-1} of amorphadiene, Y285 produced significantly less sesquiterpene: 3.3 g l^{-1} of artemisinic acid (Fig. 2a and Supplementary Table 2) plus 0.3 g l^{-1} amorphadiene, 0.18 g l^{-1} artemisinic alcohol and no detectable artemisinic aldehyde (Supplementary Table 3). The viability of the Y285 culture also decreased markedly after *CYP71AV1* and *CPR1* expression (Fig. 2a). We surmised that the decreased viability and reduced production of sesquiterpene products in Y285 might be caused by the cytochrome P450 responsible for oxidizing amorphadiene, or by the rapid accumulation of artemisinic acid.

Poor coupling between P450 cytochromes and their reductases can result in the release of reactive oxygen species⁷. In liver microsomes, the P450 enzyme is generally present in excess over its reductase⁸, whereas in Y285 and Y301 both enzymes are expressed from strong galactose-regulated promoters on a high-copy plasmid, and are presumably present at similar levels. We reduced expression of *CPR1* by expressing it from a weaker promoter (*GAL3* promoter) and integrating a single copy into genomic DNA, generating strain Y657 (Supplementary Table 1). Y657 had an increase in cell growth (Fig. 2b) and viability (Supplementary Fig. 3) compared to either Y285 or Y301 (isogenic to Y285, but with *P_{CTR3}-ERG9* replacing *P_{MET3}-ERG9*), but showed lower artemisinic acid production in shake-flask cultures (Fig. 2c) and mixed-feed fed-batch fermentors (Supplementary Tables 2 and 3). Comparison of all amorphadiene-derived sesquiterpenes showed that although reducing *CPR1* expression decreased artemisinic acid production, total sesquiterpene production remained relatively high, indicating that low *CPR1* levels increase cell health, but decrease the total rate of amorphadiene oxidations (Fig. 3a; compare Y301 and Y657). The reaction rate of some cytochromes P450 is enhanced by their interaction with cytochrome *b₅* as explained by several possible mechanisms^{9,10}. We identified a cytochrome *b₅* complementary DNA from *A. annua* (*CYB5*; Supplementary Fig. 4) and expressed a chromosomally integrated copy from a strong promoter

¹Amyris, Inc., 5885 Hollis Street, Suite 100, Emeryville, California 94608, USA. ²National Research Council of Canada, 110 Gymnasium Place, Saskatoon, Saskatchewan S7N 0W9, Canada. ³Department of Chemical and Biomolecular Engineering, University of California, Berkeley, California 94720, USA. ⁴Department of Bioengineering, University of California, Berkeley, California 94720, USA. ⁵Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ⁶Joint BioEnergy Institute, 5885 Hollis Street, Emeryville, California 94608, USA. [†]Present addresses: Dow AgroSciences, 9330 Zionsville Road, Indianapolis, Indiana 46268, USA (P.J.W.); Radiant Genomics, Inc., 2430 Fifth Street, Suite D, Berkeley, California 94710, USA (A.M.); Arkansas Biosciences Institute, Arkansas State University, Jonesboro, Arkansas 72401, USA (K.H.T.); Solazyme, Inc., 225 Gateway Boulevard, South San Francisco, California 94080, USA (J. Lenihan); Sutro Biopharma Inc., 310 Utah Avenue, South San Francisco, California 94080, USA (S.B.); Abbvie, 1500 Seaport Boulevard, Redwood City, California 94063, USA (G. Dang); Horticultural Sciences Department, University of Florida, PO Box 110690, Gainesville, Florida 32611-0690, USA (K.W.E.); Department of Plant Sciences, University of Saskatchewan, 51 Campus Drive, Saskatoon, Saskatchewan S7N 5A8, Canada (M.H.); Regentin Consulting, 25409 Modoc Court, Hayward, California 94542, USA (R.R.); AS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, The Chinese Academy of Science, Wuhan 430074, China (Y.Z.); Genomatica, Inc., 10520 Wateridge Circle, San Diego, California 92121, USA (J. Lievens).

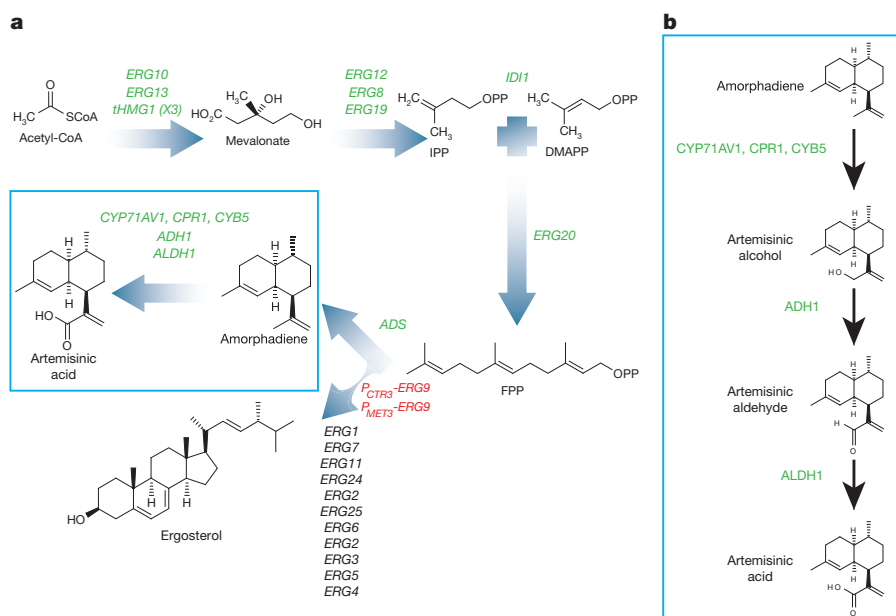


Figure 1 | Artemisinin acid production pathway in *S. cerevisiae* and summary of strains described.

a, Overview of artemisinin acid production pathway. Overexpressed genes controlled by the GAL induction system are shown in green. Copper- or methionine-repressed squalene synthase (*ERG9*) is shown in red. DMAPP, dimethylallyl diphosphate; FPP, farnesyl diphosphate; IPP, isopentenyl diphosphate. *tHMG1* encodes truncated HMG-CoA reductase. **b**, The full three-step oxidation of amorphadiene to artemisinic acid from *A. annua* expressed in *S. cerevisiae*. CYP71AV1, CPR1 and CYB5 oxidize amorphadiene to artemisinic alcohol; ADH1 oxidizes artemisinic alcohol to artemisinic aldehyde; ALDH1 oxidizes artemisinic aldehyde to artemisinic acid. Strains containing these genes are described in Supplementary Table 1.

(*GAL7* promoter) in a strain with low CPR1 expression. The resulting strain, Y692 (Supplementary Table 1), produced higher concentrations of artemisinic acid than strains without CYB5 (Fig. 2c and Supplementary Tables 2 and 3; compare Y657 and Y692). Expression of CYB5 also increased the production of artemisinic aldehyde, leading to a 40% increase in total sesquiterpene production in shake-flask cultures (Fig. 3a; compare Y657 and Y692), and almost doubled the production of artemisinic aldehyde in fermentors (Supplementary Table 3). In view of the reactivity and presumed toxicity of artemisinic aldehyde, we expressed a recently isolated cDNA encoding *A. annua* artemisinic aldehyde dehydrogenase (*ALDH1*)¹¹ in Y692 to produce Y973 and Y1368 (also expressing increased levels of cytosolic catalase to reduce oxidative stress; Supplementary Table 1). Expression of *ALDH1* markedly increased the production of artemisinic acid in both flask (Fig. 2c; Y1368) and fermentor cultures (Supplementary Tables 2 and 3; Y973). Artemisinic aldehyde was undetectable in flask cultures (Fig. 3a; Y1368), and barely detectable in fermentors (Supplementary Table 3; Y973). Furthermore, the expression of *ALDH1* in Y973 allowed early induction of fermentor cultures immediately after inoculation (previous attempts at early induction with Y285 and Y301 had resulted in rapid loss of viability), further increasing production to 7.7 g l⁻¹ artemisinic acid (Supplementary Tables 2 and 4). The yield (Cmol% of substrate carbon incorporated into artemisinic acid) was more than doubled compared to the initial Y285 cultures (Supplementary Tables 3 and 4).

In the course of investigating the biosynthesis of artemisinin in *A. annua* glandular trichomes, a gene encoding a putative alcohol dehydrogenase (*ADH1*) was examined. The gene is represented by a contiguous set of glandular trichome-derived expressed sequence tags (ESTs)¹² corresponding to 1.3% of the EST collection. The *A. annua* *ADH1* open reading frame (ORF) was expressed as a fusion protein and purified from *Escherichia coli*. Sequence analysis and *in vitro* characterization revealed that ADH1 is an NAD-dependent alcohol dehydrogenase of the medium chain dehydrogenase/reductase superfamily, with specificity towards artemisinic alcohol (Michaelis constant (K_m) = 11 ± 3 μM, k_{cat} = 41 ± 5 s⁻¹ (mean ± s.e.m.); Supplementary Fig. 5). This specificity and the evidence for strong glandular trichome expression indicate a role for ADH1 in the formation of artemisinic aldehyde in the artemisinin pathway of *A. annua*. Therefore, we propose that all five enzymes (CYP71AV1, CPR1, CYB5, ADH1 and ALDH1) are involved in the oxidation of amorphadiene to artemisinic acid in *A. annua* plants, and set out to reconstitute the entire heterologous biosynthetic pathway in yeast (Fig. 1b).

Observing the accumulation of artemisinic alcohol in strain Y1368 (Fig. 3a), we completed the biosynthetic pathway by expressing *ADH1* in conjunction with *ALDH1*, CYP71AV1, CYB5 and CPR1. The resulting strain, Y1283, produced no detectable artemisinic alcohol in flask cultures, and increased artemisinic acid production by 18% (Figs 2c and 3a; compare Y1368 and Y1283). In fermentors, *ADH1* increased production to 8.1 g l⁻¹ in an early induction, mixed-feed process, while

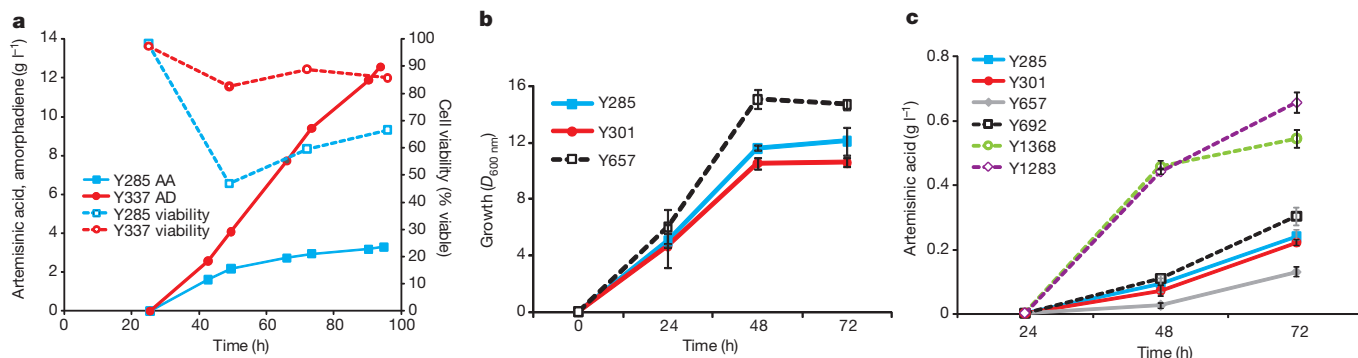


Figure 2 | Growth, viability and production by *S. cerevisiae* strains.

a, Production and cell viability of artemisinic acid production strain Y285 and amorphadiene production strain Y337 in the glucose and ethanol mixed-feed, fed-batch fermentation process. AA, artemisinic acid; AD, amorphadiene.

b, Growth of artemisinic-acid-producing strains in shake-flasks. **c**, Production of artemisinic acid in shake-flasks by different strains. Error bars denote standard deviation of triplicate shake-flask cultures.

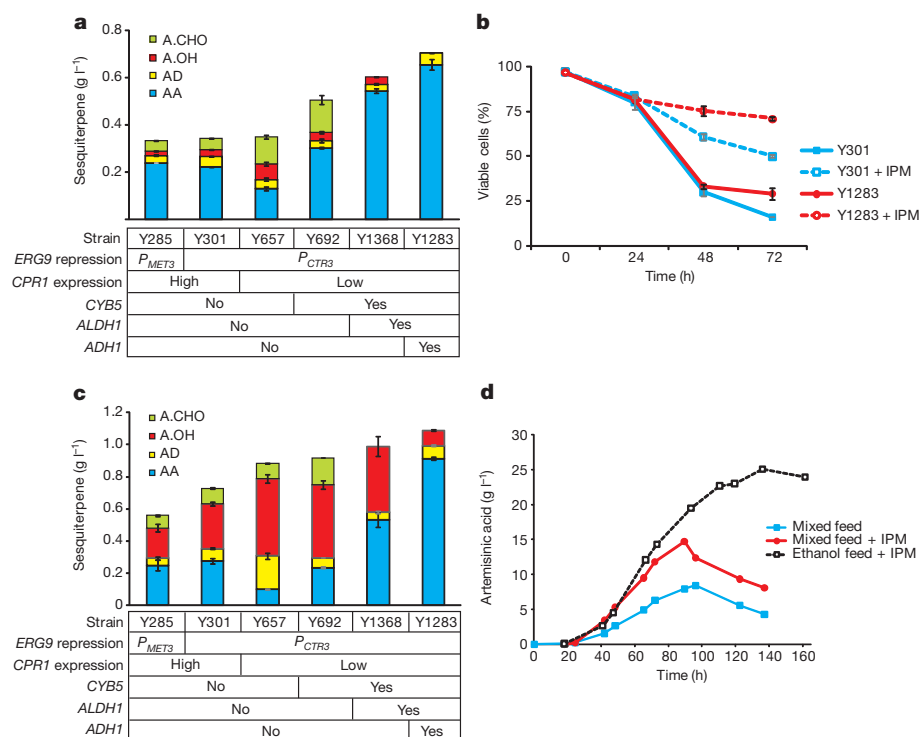


Figure 3 | Increasing production of artemisinic acid by strain engineering and addition of IPM to cultures. **a**, Artemisinic acid production in flasks (no IPM). A.CHO, artemisinic aldehyde; A.OH, artemisinic alcohol. **b**, Effect of IPM addition on the viability of strains Y301 and Y1283 in shake-flask cultures. **c**, Production of artemisinic acid in flasks containing IPM. **d**, Artemisinic acid production of strain Y1284 in fed-batch fermentation processes. All fermentations were run with early repression (150 μ M CuSO_4 was added before inoculation). Because Y1284 has the *gal80 Δ* genotype, no galactose was added to the fermentations. Error bars denote standard deviation of triplicate shake-flask cultures.

reducing production of artemisinic alcohol (Supplementary Table 4). The improved viability (Supplementary Fig. 6) of Y1283, which also contained the added catalase activity, allowed us to express all heterologous galactose-regulated enzymes constitutively by deletion of *GAL80*, a strategy used previously to increase production in amorphadiene-production strains³. This strain, Y1284, does not require the inducer galactose, yet produces higher concentrations of artemisinic acid compared to its parent (Supplementary Tables 2–4).

We observed that strains expressing ALDH1 produced artemisinic acid as a crystalline extracellular precipitate (Supplementary Figs 7 and 8). Precipitation was also observed in early induction fed-batch mixed-feed fermentors, complicating accurate measurement of product from the heterogeneous fermentor samples. To overcome the sampling difficulties associated with artemisinic acid precipitation, we investigated the effect of solubilizing the precipitate by extractive fermentation¹³, growing the cultures in the presence of isopropyl myristate (IPM) oil. Addition of 10% (v/v) IPM to flask cultures resulted in a marked increase in the viability of all strains, shown for Y301 and Y1283 in Fig. 3b. Whereas the addition of IPM to earlier strains (Y285, Y301, Y657 and Y692) lacking ALDH1 and ADH1 resulted in extraction of intermediates (amorphadiene, artemisinic alcohol and aldehyde; Fig. 3c), IPM addition to strains containing ALDH1 and ADH1 produced artemisinic acid to $>14 \text{ g l}^{-1}$ in a mixed-feed, early induction fermentation process (Supplementary Table 2), with a six-fold increase in yield compared to Y285 (Supplementary Table 5). The IPM extractive fermentation of Y1284 allowed us to increase production further by developing a feedback-controlled ethanol pulse-feed process; in this process Y1284 produced 25 g l^{-1} of artemisinic acid, with a sevenfold higher yield than Y285 (Fig. 3d and Supplementary Tables 2 and 5). Artemisinic acid production in the mixed-feed process was seen to reach a maximum concentration due to precipitation from the culture broth mixture and a cessation in production while the fed-batch fermentation continued. Additional IPM was added to the ethanol-feed fermentation to avoid precipitation. To take advantage of this high-titre process we developed a method for extracting artemisinic acid from IPM with high yield and purity (see Methods).

Numerous total and partial chemical syntheses of artemisinin exist^{14–16}. Here we report one starting from our purified artemisinic acid (Fig. 4a). For this route to be scalable and practical, several modifications of the previously published syntheses are required. The first step is the reduction of the $\Delta^{11}(13)$ double bond (numbering system of Sy and Brown¹⁷) to give dihydroartemisinic acid, which has two epimers, of which only the (*R*)-11 one has the correct stereochemistry found in artemisinin. This had typically been carried out with ‘nickel boride’ (NaBH_4 or LiBH_4 plus NiCl_2), which gives a $\sim 3:1$ to $85:15$ ratio of the 11-epimers, favouring the desired one, but the use of a stoichiometric excess of the reducing agent and the poor isomer ratio are unacceptable for a cost-effective scalable synthesis. We found that catalytic hydrogenation using several different noble metal catalysts affords nearly quantitative yields of the reduced acid in (*R*)-11:(*S*)-11 epimer ratios as high as 94:6, without any significant overreduction to tetrahydroartemisinic acid (Supplementary Fig. 9 and Supplementary Table 6).

The next step is the esterification of the carboxylic acid. The subsequent reactions can be performed using the acid ($\text{R} = \text{H}$), but this results in considerable yield losses owing to the formation of the five-membered lactone-containing compound dihydroepideoxyartemisinin B, a side-reaction blocked by the presence of the ester. Large-scale esterification was readily accomplished by carboxyl activation by acid chloride formation, followed by an alcohol quench (Supplementary Fig. 10).

The third step is an ‘ene-type’ reaction of the C4–C5 double bond with singlet oxygen ($^1\text{O}_2$) to give an allylic 3-hydroperoxide. In previous syntheses, the $^1\text{O}_2$ was invariably generated by photosensitized energy transfer from dye molecules (such as rose bengal, methylene blue and porphyrins)^{18–20}, but because photosynthetic steps are rarely found in manufacturing facilities, we sought another source of $^1\text{O}_2$. A practical alternative was found in the group VI metal salt-induced disproportionation of concentrated H_2O_2 (ref. 21). With this technique the hydroperoxide could be formed cleanly with no evidence of isomers or rearrangement products.

In the final step the allylic hydroperoxide undergoes an acid-catalysed Hock fragmentation and rearrangement to afford a ring-opened keto-aldehyde enol. Trapping of this enol with $^3\text{O}_2$ produces

a vicinal hydroperoxide aldehyde, which in a cascade of acid-catalysed cyclizations forms an endoperoxide bridge, a seven-membered cyclic ether, and finally a six-membered lactone, thus producing good yields of artemisinin with the correct stereochemistry (Supplementary Figs 11–13). Improvements of this step involved substitution of expensive copper triflate¹⁶ with benzenesulphonic acid/sulphonate copper(II) DOWEX resin and replacement (for safety reasons) of pure O₂ by air, in addition to reaction condition optimization to maximize the yield of artemisinin. These changes produced a scalable four-step synthesis that gave artemisinin in 40–45% overall yield, a marked

improvement over the typical yields reported in the literature^{18–20} (Supplementary Table 7). The product is purer than the plant-sourced artemisinin used in ACT production, as shown by high-performance liquid chromatography (HPLC) comparison of our material with a selection of commercial plant samples (Fig. 4b and Supplementary Table 8).

Our results describe for the first time, to our knowledge, the expression of the complete pathway for artemisinic acid production, which resulted in a greater than tenfold increase in artemisinic acid titres. In addition, we demonstrated a significant increase in the efficiency of artemisinic acid conversion to artemisinin compared with earlier work^{18–20}. We show that expression of CYP71AV1 and its cognate reductase is not sufficient for high-level production of artemisinic acid, instead requiring three additional plant enzymes (CYB5, ADH1 and ALDH1 (ref. 11); Fig. 1b). Optimization of the CYP71AV1:CPRI expression ratio, combined with CYB5 expression, overcame initial viability problems, but high titre and increased yield were only achieved when *A. annua* artemisinic alcohol and aldehyde dehydrogenases were co-expressed. These observations in yeast lend strong support to the importance of the dehydrogenases in artemisinin biosynthesis in the native plant system²². Extractive fermentation with a bio-compatible solvent (IPM) increased the production of artemisinic acid through an uncharacterized effect, perhaps by effectively removing the acid from the aqueous phase. The development of a facile procedure for purification of artemisinic acid from IPM in high yield and purity allows a ready supply for chemical conversion to artemisinin. The chemical conversion procedure is notable for its simplicity, scalability, economy of reagents, and the high yield obtained. Nevertheless, given investment in suitable large-scale photoreactors, improvements of the classical photochemistry syntheses have the potential to increase the overall process yields further (see Supplementary Information for details and comparison to photochemical syntheses^{18–20}).

In summary, we have determined the full artemisinic acid biosynthetic pathway and developed a process for the production of the antimalarial drug artemisinin by fermentation of simple inexpensive carbon substrates using engineered *S. cerevisiae* to produce artemisinic acid, followed by extraction and chemical conversion to artemisinin. These key developments in yeast strain engineering, fermentation, and artemisinin synthetic chemistry pave the way for an industrial process capable of supplementing the world supply of artemisinin from a second source independent of the uncertainties associated with botanical production.

METHODS SUMMARY

The *A. annua* cytochrome *b*₅ cDNA sequence was identified from a trichome expressed sequence tag library (NCBI accession 35608) by searching for sequence similarity to *Crepis alpina* cytochrome *b*₅ type 11 (ref. 23). A cDNA encoding *A. annua* ADH1 was identified and the encoded protein characterized essentially as described for ALDH1 (ref. 11). Growth of strains, general genetic methodology and construction of synthetic genes were as described³. Yeast strains were derived from Y337 (ref. 3). Salient features of DNA constructs are as follows: (1) for expression of *ERG9* from the *CTR3* promoter the *MET3* promoter was replaced with nucleotides –1 to –734 of the *CTR3* promoter, integration being selected by D-serine²⁴ (*erg9Δ::dsdA_P_{CTR3}-ERG9*) or nourseothricin²⁵ (*erg9Δ::natA_P_{CTR3}-ERG9*) resistance; (2) for reduced expression of cytochrome P450 reductase (*CPRI*), *CPRI* was removed from plasmid pAM322 (ref. 3) by digestion and recircularization to generate pAM552, which expresses only ADS and CYP71AV1. A single copy of *CPRI* was expressed from the *GAL3* promoter (*P_{GAL3}*) (nucleotides –1 to –660) integrated between *GAL1* and *GAL7* (*gal1/10/7Δ::natA_P_{GAL3}-CPRI-T_{CYC1}*, in which *T_{CYC1}* denotes the *CYC1* terminator); (3) a single integrated copy of *A. annua* cytochrome *b*₅ was expressed from the *GAL7* promoter (nucleotides –1 to –725; *leu2::hisMXA::kanA_P_{GAL7}-CYB5-T_{CYC1}*); (4) a single integrated copy of *A. annua* aldehyde dehydrogenase (*ALDH1*) was expressed from the *GAL7* promoter, selecting for hygromycin B (ref. 25) (*ndt80::hphA_P_{GAL7}-ALDH1-T_{TDH1}* and *his3::hphA_P_{GAL7}-ALDH1-T_{TDH1}*); (5) a single integrated copy of *A. annua* alcohol dehydrogenase (*ADH1*) was expressed from the *GAL7* promoter, selecting for uracil prototrophy (*natAΔ::URA3_P_{GAL7}-ADH1-T_{TDH1}* and *gal80Δ::URA3_P_{GAL7}-ADH1-T_{GAL80}*).

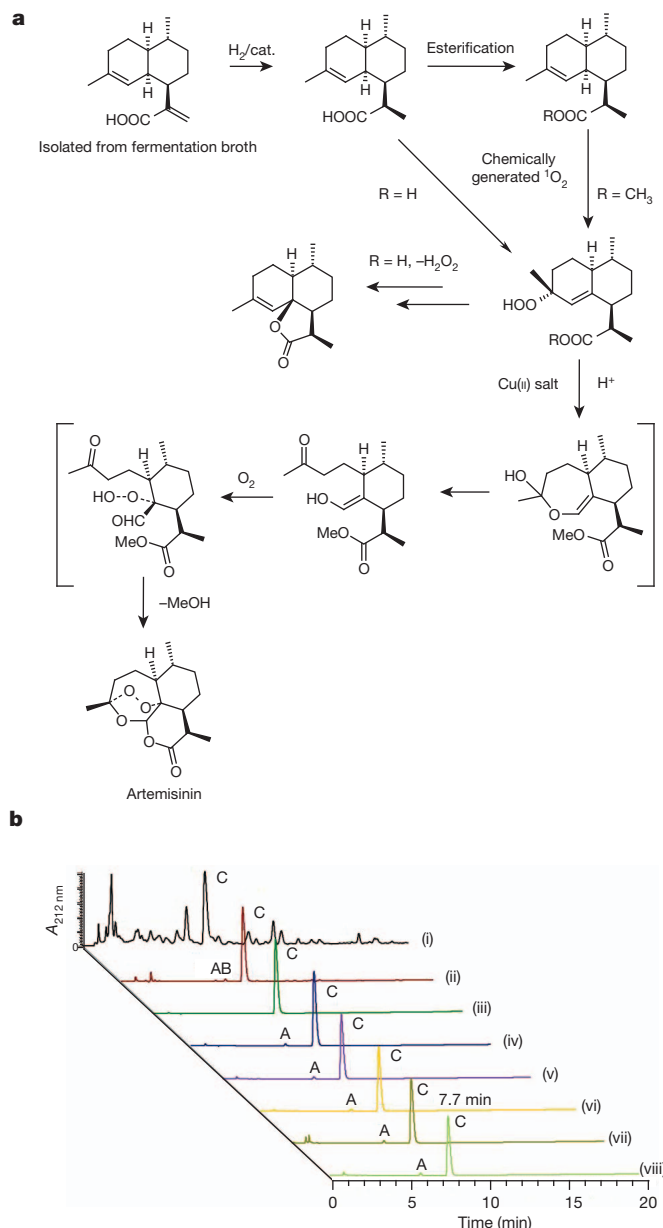


Figure 4 | Chemical conversion of artemisinic acid to artemisinin. **a**, Semi-synthesis of artemisinin from microbially produced artemisinic acid. Compounds in square brackets denote intermediates detected but not isolated. **b**, HPLC ultraviolet traces of crude, partially purified and purified semi-synthetic artemisinin and commercial samples of plant-derived artemisinin. For HPLC conditions see Supplementary Information. Major peak identification (retention time in min): artemisitene (5.93–5.96, peak A), 9-*epi*-artemisinin (6.54, peak B); artemisinin (7.72–7.74, peak C). Traces: (i) crude reaction mixture: <33% pure; (ii) crude mixture filtered through silica gel: 90.5% pure; (iii) recrystallized: 99.6% pure; (iv–viii) samples of plant artemisinin from different Chinese commercial suppliers, all quality control approved as raw materials suitable for ACT manufacture (97.2–97.8% pure).

Flask and fermentor culture conditions were essentially as described³. Fermentations requiring IPM contained 400 ml IPM added to 800 ml fermentor volume before inoculation. Artemisinic acid was purified from IPM by aqueous extraction at pH 10.7, followed by precipitation at pH 5.0. Assays for amorph-4,11-diene and artemisinic acid are essentially as described³. Artemisinic alcohol and artemisinic aldehyde were monitored by gas chromatography with flame-ionization detection.

Full Methods and any associated references are available in the online version of the paper.

Received 8 July 2012; accepted 4 March 2013.

Published online 10 April; corrected online 24 April 2013 (see full-text HTML version for details).

- World Health Organization. *World Malaria Report 2011* <http://www.who.int/malaria/publications/atoz/9789241564403/en/index.html> (World Health Organization, 2011).
- Hale, V., Keasling, J. D., Renninger, N. & Diagana, T. T. Microbially derived artemisinin: a biotechnology solution to the global problem of access to affordable antimalarial drugs. *Am. J. Trop. Med. Hyg.* **77**, 198–202 (2007).
- Westfall, P. J. *et al.* Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc. Natl Acad. Sci. USA* **109**, E111–E118 (2012).
- Ro, D. K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).
- Paradise, E. M., Kirby, J., Chan, R. & Keasling, J. D. Redirection of flux through the FPP branch-point in *Saccharomyces cerevisiae* by down-regulating squalene synthase. *Biotechnol. Bioeng.* **100**, 371–378 (2008).
- Asadollahi, M. A. *et al.* Production of plant sesquiterpenes in *Saccharomyces cerevisiae*: effect of ERG9 repression on sesquiterpene biosynthesis. *Biotechnol. Bioeng.* **99**, 666–677 (2008).
- Zangar, R. C., Davydov, D. R. & Verma, S. Mechanisms that regulate production of reactive oxygen species by cytochrome P450. *Toxicol. Appl. Pharmacol.* **199**, 316–331 (2004).
- Peterson, J. A., Ebel, R. E., O'Keeffe, D. H., Matsubara, T. & Estabrook, R. W. Temperature dependence of cytochrome P-450 reduction. A model for NADPH-cytochrome P-450 reductase:cytochrome P-450 interaction. *J. Biol. Chem.* **251**, 4010–4016 (1976).
- Schenkman, J. B. & Jansson, I. The many roles of cytochrome *b*₅. *Pharmacol. Ther.* **97**, 139–152 (2003).
- Zhang, H., Im, S. C. & Waskell, L. Cytochrome *b*₅ increases the rate of product formation by cytochrome P450 2B4 and competes with cytochrome P450 reductase for a binding site on cytochrome P450 2B4. *J. Biol. Chem.* **282**, 29766–29776 (2007).
- Teoh, K. H., Polichuk, D. R., Reed, D. W. & Covello, P. S. Molecular cloning of an aldehyde dehydrogenase implicated in artemisinin biosynthesis in *Artemisia annua*. *Botany* **87**, 635–642 (2009).
- Teoh, K. H., Polichuk, D. R., Reed, D. W., Nowak, G. & Covello, P. S. *Artemisia annua* L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. *FEBS Lett.* **580**, 1411–1416 (2006).
- Daugulis, A. J. Partitioning bioreactors. *Curr. Opin. Biotechnol.* **8**, 169–174 (1997).
- Li, Y., Huang, H. & Wu, Y.-L. in *Medicinal Chemistry of Bioactive Natural Products* (eds Liang, X.-T. & Fang, W.-S.) 183–256 (John Wiley and Sons, 2006).
- Kim, B. & Sasaki, T. Recent progress in the synthesis of artemisinin and its derivatives. *Org. Prep. Proced. Int.* **38**, 1–80 (2006).
- Haynes, R. K. From artemisinin to new artemisinin antimalarials: biosynthesis, extraction, old and new derivatives, stereochemistry and medicinal chemistry requirements. *Curr. Top. Med. Chem.* **6**, 509–537 (2006).
- Sy, L.-K. & Brown, G. D. The role of the 12-carboxylic acid group in the spontaneous autoxidation of dihydroartemisinic acid. *Tetrahedron* **58**, 909–923 (2002).
- Haynes, R. K. & Vonwiller, S. C. Catalysed oxygenation of allylic hydroperoxides derived from qinghao (artemisinic) acid. Conversion of qinghao acid into dehydroqinghaosu (artemisitene) and qinghaosu (artemisinin). *J. Chem. Soc. Chem. Commun.* **6**, 451–453 (1990).
- Roth, R. J. & Roth, N. A. Simple conversion of artemisinic acid into artemisinin. US patent 4,992 561 (1991).
- Lévesque, F. & Seeburger, P. H. Continuous-flow synthesis of the anti-malaria drug artemisinin. *Angew. Chem. Int. Edn Engl.* **51**, 1706–1709 (2012).
- Boehme, K. & Brauer, H.-D. Generation of singlet oxygen from hydrogen peroxide disproportionation catalyzed by molybdate ions. *Inorg. Chem.* **31**, 3468–3471 (1992).
- Bertea, C. M. *et al.* Identification of intermediates and enzymes involved in the early steps of artemisinin biosynthesis in *Artemisia annua*. *Planta Med.* **71**, 40–47 (2005).
- Nam, J. W. & Kappock, T. J. Cloning and transcriptional analysis of *Crepis alpina* fatty acid desaturases affecting the biosynthesis of crepenynic acid. *J. Exp. Bot.* **58**, 1421–1432 (2007).
- Vorachek-Warren, M. K. & McCusker, J. H. DsdA (D-serine deaminase): a new heterologous MX cassette for gene disruption and selection in *Saccharomyces cerevisiae*. *Yeast* **21**, 163–171 (2004).
- Goldstein, A. L. & McCusker, J. H. Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast* **15**, 1541–1553 (1999).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Rathbone for advice on native *A. annua* ADH1 expression, and our friends and colleagues at Sanofi, especially D. Thibaut, C. Lehmann, C. Masson-Brocard, B. Dumas, P. Baduel and H. Farret. We also thank J. Rine, P. Ortiz de Montellano and H. van Dijken for many conversations. This research was conducted under the sponsorship of the Institute for OneWorld Health through generous support of the Bill & Melinda Gates Foundation for this non-profit project.

Author Contributions P.J.W., D.J.P., K.B., K.F., A.T., A.M., D.E., D.R.P., K.H.T., D.W.R., T.T., J. Lenihan, M.F., S.B., G. Dang, D. Dengrove, D. Diola, G. Dorin, K.W.E., S.F., J.G., S.P.G., T.G., R.H., M.H., T.H., T.I., H.J., L.K., B.L., D.M., N.M., S.S., H.T., R.V., L.F.W., L.X., M.Y., Y.Z. and L.Z. performed experiments; C.J.P., P.J.W., D.J.P., K.B., K.F., D.M., M.D.L., T.T., J.L., S.B., S.F., J.G., S.P.G., L.K., R.R., L.Z., J. Lievense, P.S.C., J.D.K., K.K.R., N.S.R. and J.D.N. designed experiments and analysed data; C.J.P., D.M. and J.D.N. supervised and coordinated experiments; C.J.P. wrote the paper. C.J.P., P.J.W., D.J.P. and K.B. contributed equally to this work.

Author Information Novel DNA sequences have been deposited in GenBank with the following accession numbers: native *A. annua* ADH1, JF910157; codon-optimized *A. annua* ADH1, JQ582842; codon-optimized *A. annua* CYB5, JQ582841; codon-optimized *A. annua* ALDH1, JQ609276. Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.J.P. (paddon@amyris.com) or J.D.N. (newman@amyris.com).

METHODS

The *A. annua* cytochrome *b₅* cDNA sequence was identified from a trichome expressed sequence tag library (NCBI accession 35608) by searching for sequence similarity to *Crepis alpina* cytochrome *b₅* type 11 (ref. 23). Dominant selection markers for yeast strain engineering were D-serine²⁴, nourseothricin²⁵ and hygromycin B²⁵.

Yeast strain engineering. *S. cerevisiae* codon-optimized synthetic genes of *A. annua* *ADS*, *CYP71AV1* and *CPRI* have been previously described³. Codon-optimized synthetic genes for *A. annua* *CYB5* (GenBank accession JQ582841), *A. annua* *ALDH1* (JQ609276) and *A. annua* *ADH1* (JQ582842) were synthesized by DNA 2.0 (<https://www.dna20.com/>) or Biosearch Technologies.

Construction of genome integration cassettes. The oligonucleotide primers used in this study are listed in Supplementary Table 9.

***dsdA-P_{CTR3}-ERG9*.** Replacement of the *MET3* promoter with the *CTR3* promoter in Y301 and Y592 was accomplished as follows. The *dsdA* gene (encoding D-serine deaminase) was amplified from pAM577 (containing the promoter and terminator of *Kluyveromyces lactis* *TEF1*) by PCR amplification with oligonucleotides PW91-031-CPK275-G and DE_PW91-027-CPK262-G. PCR amplification of the wild-type *CTR3* promoter from positions -1 to -734 was performed with oligonucleotides PW61-104-CPK116-G and DE_PW91-027-CPK263-G using CEN.PK2-1C (ref. 3) genomic DNA as the template. These two PCR products shared a 44-base-pair (bp) overlap at the 3' end of the promoter and the 5' end of the gene. For the secondary PCR, 25 ng each of the purified *CTR3* promoter (-1 to -734) and *dsdA* PCR were used as the DNA templates and PCR amplified with oligonucleotides PW91-031-CPK275-G and PW61-104-CPK116-G to give *P_{CTR3}(-1 to -734)-dsdA*.

***gal1/10/7::natA-P_{GAL3}-CPRI-T_{CYC1}*.** Targeted integration of the *P_{GAL3}-CPRI* expression cassette in Y657 at the *GAL7* locus was accomplished as follows. PCR amplification of the wild-type *GAL7* locus from positions 30 to 1021 was performed with oligonucleotides PW91-014-CPK236-G and PW-91-079-CPK384-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *CPRI* ORF and *CYC1* terminator was performed with oligonucleotides PW-91-079-CPK385-G and PW-91-079-CPK392-G using plasmid pAM322 (ref. 3) as the template. PCR of the wild-type *GAL3* promoter from positions -1 to -660 was performed with oligonucleotides PW-91-079-CPK393-G and PW-91-079-CPK394-G using CEN.PK2-1C genomic DNA as the template. PCR of the *natA* marker (nourseothricin resistance) was performed with oligonucleotides PW-91-079-CPK383-G and PW-91-079-CPK395-G. Each of the DNA elements from the first round of PCR was designed to share a 20–30-bp overlap with the adjacent element, by using non-templated tails on the oligonucleotides. For the secondary PCR amplification, 25 ng each of the purified *GAL7*, *CPRI-CYC1*, *GAL3* promoter and *natA* PCR products were used as the DNA template and PCR amplified with oligonucleotides PW91-014-CPK236-G and PW-91-079-CPK383-G to give *GAL7(30 to 1021)-P_{GAL3}(-1 to -660)-CPRI-T_{CYC1}-natA*.

***leu2::hisMXA::kanA-P_{GAL7}-CYB5-T_{CYC1}*.** Targeted replacement of the *leu2::hisMX* locus in Y657 was accomplished as follows. PCR amplification of the wild-type *ERG19* locus from positions 489 to 1341 was performed with oligonucleotides AM/PW-91-093-CPK461-G and AM/PW-91-093-CPK462-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *kanA* marker (G418 resistance) was performed with oligonucleotides AM/PW-91-093-CPK460-G and AM-125-50-CPK514-G using pAM575 (containing the promoter and terminator of *K. lactis* *TEF1*) as the template. PCR amplification of the *GAL7* promoter from positions -1 to -725 was performed with oligonucleotides AM-125-50-CPK513-G and AT-126-103-CPK593-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *S. cerevisiae* codon-optimized *A. annua* *CYB5* ORF was performed with oligonucleotides AT-126-103-CPK592-G and PW-91-093-CPK426-G. PCR amplification of the *CYC1* terminator (*T_{CYC1}*) from positions 331 to 830 was performed with oligonucleotides PW-91-093-CPK425-G and AT-126-103-CPK595-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *LEU2* locus from positions 1 to 450 was performed with oligonucleotides AT-126-103-CPK594-G and AM/PW-91-093-CPK457-G using CEN.PK2-1C genomic DNA as the template. For the secondary PCR, 25 ng each of the purified *ERG19*(489 to 1341), *kanA*, *P_{GAL7}(-1 to -725)*, *A. annua* *CYB5*, *T_{CYC1}(331 to 830)* and *LEU2*(1 to 450) PCR products were used as the DNA template and PCR amplified with oligonucleotides AM/PW-91-093-CPK462-G and AM/PW-91-093-CPK457-G to give *ERG19*(489 to 1341)-*kanA-P_{GAL7}(-1 to -725)-CYB5-T_{CYC1}(331 to 830)-LEU2*(1 to 450).

***ndt80Δ::P_{TDH1}-HEM1-hphA-P_{PGK1}-CTT1*.** Targeted replacement of the *NDT80* locus in Y1368 with constitutively expressed *CTT1* and *HEM1* was accomplished as follows. PCR amplification of the wild-type *NDT80* locus from positions -187 to -951 was performed with oligonucleotides PW-091-144-CPK640-G and PW-091-144-CPK654-G using CEN.PK2-1C genomic DNA as

the template. PCR of the wild-type *HEM1* locus from positions 1 to 1947 was performed with oligonucleotides PW-091-144-CPK655-G and PW-091-144-CPK656-G using CEN.PK2-1C genomic DNA as the template. PCR of the *TDH1* promoter (*P_{TDH1}*) from positions -1 to -577 was performed with oligonucleotides PW-091-144-CPK657-G and PW-091-144-CPK658-G using CEN.PK2-1C genomic DNA as the template. PCR of the *hphA* marker was performed with oligonucleotides PW-091-144-CPK659-G and PW-091-144-CPK643-G using BY4710 (ref. 4) genomic DNA as the template. PCR of the *PGK1* promoter (*P_{PGK1}*) from positions -1 to -623 was performed with oligonucleotides PW-091-144-CPK644-G and PW-091-144-CPK645-G using CEN.PK2-1C genomic DNA as the template. PCR of the *CTT1* locus from positions 1 to 2000 was performed with oligonucleotides PW-091-144-CPK646-G and PW-091-144-CPK647-G using CEN.PK2-1C genomic DNA as the template. PCR of the wild-type *NDT80* locus from positions 1684 to 2470 was performed with oligonucleotides PW-091-144-CPK648-G and PW-091-144-CPK649-G using CEN.PK2-1C genomic DNA as the template. For the secondary PCR, 25 ng each of the purified *NDT80*(-187 to -951), *HEM1*(1 to 1947), *P_{TDH1}(-1 to -577)*, *hphA*, *P_{PGK1}(-1 to -623)*, *CTT1*(1 to 2000) and *NDT80*(1684 to 2470) PCR products were used as the DNA template and PCR amplified with oligonucleotides PW-091-144-CPK640-G and PW-091-144-CPK649-G to give *NDT80*(-187 to -951)-*P_{TDH1}(-1 to -577)-HEM1*(1 to 1947)-*hphA-P_{PGK1}(-1 to -623)-CTT1*(1 to 2000)-*NDT80*(1684 to 2470).

***ndt80::hphA-P_{GAL7}-ALDH1-T_{TDH1}*.** Targeted replacement of the *NDT80* locus with *P_{GAL7}-ALDH1* in Y973 was accomplished as follows. PCR amplification of the wild-type *NDT80* locus from positions -187 to -951 was performed with oligonucleotides PW-091-144-CPK640-G and PW-091-144-CPK641-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *hphA* marker was performed with oligonucleotides PW-091-144-CPK642-G and AM-125-50-CPK514-G using pAM578 pAM575 (containing the promoter and terminator of *K. lactis* *TEF1*) as the template. PCR amplification of the *GAL7* promoter (*P_{GAL7}*) from positions -1 to -725 was performed with oligonucleotides AM-125-50-CPK513-G and AM-125-107-CPK756-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *S. cerevisiae* codon-optimized *ALDH1* ORF was performed with oligonucleotides AM-125-107-CPK754-G and AM-125-107-CPK755-G. PCR amplification of the *TDH1* terminator (*T_{TDH1}*) from positions 1000 to 1997 was performed with oligonucleotides AM-125-107-752G and AM-125-107-CPK753-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the wild-type *NDT80* locus from positions 1684 to 2470 was performed with oligonucleotides AM-125-107-CPK751-G and PW-091-144-CPK649-G using CEN.PK2-1C genomic DNA as the template. For the secondary PCR amplification, 25 ng each of the purified *NDT80*(-187 to -951), *hphA*, *P_{GAL7}(-1 to -725)*, *A. annua* *ALDH1*, *T_{TDH1}(1000 to 1997)* and *NDT80*(1684 to 2470) PCR products were used as the DNA templates and PCR amplified with oligonucleotides PW-091-144-CPK640-G and PW-091-144-CPK649-G to give *NDT80*(-187 to -951)-*hphA-P_{GAL7}(-1 to -725)-ALDH1-T_{TDH1}(1000 to 1997)-NDT80*(1684 to 2470).

***his3::hphA-P_{GAL7}-ALDH1-T_{TDH1}*.** Targeted replacement of the *his3::hisMX* locus with *P_{GAL7}-ALDH1* to create Y1368 was accomplished as follows. PCR amplification of the wild-type *HIS3* locus from positions -32 to -630 was performed with oligonucleotides PW-91-129-CPK543-G and PW-91-129-CPK544-G using BY4710 genomic DNA as the template. PCR of the *hphA* marker was performed with oligonucleotides PW-91-129-CPK545-G and AM-125-50-CPK514-G using pAM578 as the template. PCR amplification of the *GAL7* promoter (*P_{GAL7}*) from positions -1 to -725 was performed with oligonucleotides AM-125-50-CPK513-G and AM-125-107-CPK756-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *A. annua* *ALDH1* ORF was performed with oligonucleotides AM-125-107-CPK754-G and AM-125-107-CPK755-G using a synthetic, *S. cerevisiae* codon-optimized template (DNA2.0). PCR amplification of the *TDH1* terminator (*T_{TDH1}*) from positions 1000 to 1997 was performed with oligonucleotides PW-191-015-CPK859-G and AM-125-107-CPK753-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the wild-type *ERG12* locus from positions 883 to 1456 was performed with oligonucleotides PW-191-015-CPK860-G and PW-91-129-CPK550-G using CEN.PK2-1C genomic DNA as the template. For the secondary PCR amplification, 25 ng each of the purified *HIS3*(-32 to -630), *hphA*, *P_{GAL7}(-1 to -725)*, *A. annua* *ALDH1*, *T_{TDH1}(1000 to 1997)* and *ERG12*(883 to 1456) PCR products were used as the DNA templates and PCR amplified with oligonucleotides PW-91-129-CPK543-G and PW-91-129-CPK550-G to give *NDT80*(-187 to -951)-*hphA-P_{GAL7}(-1 to -725)-ALDH1-T_{TDH1}(1000 to 1997)-ERG12*(883 to 1456).

***natAΔ::URA3-P_{GAL7}-ADH1-T_{TDH1}*.** Targeted replacement of the *natA* locus with *P_{GAL7}-ADH1* for creation of Y1283 was accomplished as follows. PCR amplification of the wild-type *GAL3* promoter (*P_{GAL3}*) from positions -77 to -660 was performed with oligonucleotides PW-191-015-CPK866-G and

PW-191-015-CPK867-G using BY4710 genomic DNA as the template. PCR amplification of the *URA3* locus from position -226 to 884 was performed with oligonucleotides PW-191-015-CPK868-G and PW-191-015-CPK869-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *GAL7* promoter (*P_{GAL7}*) from positions -1 to -725 was performed with oligonucleotides PW-191-015-CPK870-G and PW-191-015-CPK871-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *S. cerevisiae* codon-optimized *A. annua* *ADH1* ORF was performed with oligonucleotides PW-191-015-CPK872-G and PW-191-015-CPK873-G. PCR amplification of the *TDH1* terminator (*T_{TDH1}*) from positions 1000 to 1750 was performed with oligonucleotides PW-191-015-CPK874-G and PW-191-015-CPK875-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the wild-type *GAL1* locus from positions 1637 to 2436 was performed with oligonucleotides PW-191-015-CPK876-G and PW-191-015-CPK877-G using CEN.PK2-1C genomic DNA as the template. For the secondary PCR amplification, 25 ng each of the purified *P_{GAL3}*(-77 to -660), *URA3*(-226 to 884), *P_{GAL7}*(-1 to -725), *A. annua* *ADH1*, *T_{TDH1}*(1000 to -1750) and *GAL1*(1637 to 2436) PCR products were used as the DNA templates and PCR amplified with oligonucleotides PW-191-015-CPK866-G and PW-191-015-CPK877-G to give *P_{GAL3}*(-77 to -660)-*URA3*(-226 to 884)-*P_{GAL7}*(-1 to -725)-*ADH1*-*T_{TDH1}*(1000 to -1750)-*GAL1*(1637 to 2436).

gal80Δ::URA3-*P_{GAL7}*-*ADH1*-*T_{GAL80}*. Targeted replacement of the *GAL80* locus with *P_{GAL7}*-*ADH1* to create Y1284 was accomplished as follows. PCR amplification of the wild-type *GAL80* locus from positions -28 to -760 was performed with oligonucleotides PW-191-015-CPK882-G and PW-191-015-CPK883-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *URA3* locus from position -226 to 884 was performed with oligonucleotides PW-191-015-CPK884-G and PW-191-015-CPK869-G using BY4710 genomic DNA as the template. PCR of the *GAL7* promoter (*P_{GAL7}*) from positions -1 to -725 was performed with oligonucleotides PW-191-015-CPK870-G and PW-191-015-CPK871-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *A. annua* *ADH1* ORF was performed with oligonucleotides PW-191-015-CPK872-G and PW-191-015-CPK873-G using a synthetic *S. cerevisiae* codon-optimized template (DNA2.0). PCR amplification of the wild-type *GAL80* locus from positions 1320 to 2117 was performed with oligonucleotides PW-191-015-CPK886-G and PW-191-015-CPK887-G using CEN.PK2-1C genomic DNA as the template. For the secondary PCR amplification, 25 ng each of the purified *GAL80*(-28 to -760), *URA3*(-226 to 884), *P_{GAL7}*(-1 to -725), *A. annua* *ADH1*, and *GAL80*(1320 to 2117) PCR products were used as the DNA templates and PCR amplified with oligonucleotides PW-191-015-CPK882-G and PW-191-015-CPK887-G to give *GAL80*(-28 to -760)-*URA3*(-226 to 884)-*P_{GAL7}*(-1 to -725)-*ADH1*-*GAL80*(1320 to 2117).

All strains were confirmed with diagnostic PCR to contain the expected integration constructs and, where appropriate, all integrations were verified by sequence analysis.

Cloning and characterization of *A. annua* *ADH1*. Analysis of a previously developed *A. annua* EST collection^{12,26,27} identified a contig corresponding to an apparently full-length ORF encoding a putative trichome-expressed alcohol dehydrogenase. The corresponding gene, designated *A. annua* *ADH1*, was associated with 2.2%, 1.3% and 0.06% of ESTs in the 'trichome-minus-flower-bud' (designated GSTSUB in ref. 12), glandular trichome (designated AAGST¹²) and flower bud (designated AAFB¹²) collections, respectively. Similarly, based on the generation of expressed sequence tags by 454 sequencing of *A. annua*²⁸, the gene expression pattern of *ADH1* was found to be comparable to that of *CYP71AV1* in a range of tissues, with negligible expression in cotyledons and mature leaf trichomes and 0.21, 0.53 and 0.03% of sequences in each of the EST collections derived from young leaf trichomes, flower bud trichomes and meristem/young leaf, respectively. A full-length *ADH1* ORF was cloned by reverse transcriptase PCR (RT-PCR) (using oligonucleotide primers PSC1 and PSC2 and the vector pENTR/D TOPO (Invitrogen)). The *A. annua* *ADH1* gene has an ORF encoding a polypeptide of 378 amino acids with a relative molecular mass of 40,415 daltons. On the basis of sequence similarities, *A. annua* *ADH1* is a member of the medium chain alcohol dehydrogenase/reductase superfamily that is related to predicted proteins of *Populus trichocarpa* (61% identity, GenBank accession XP_002324694) and *Cynara cardunculus* (72% identity over 214 amino acids, GenBank accession GE588275).

The *A. annua* *ADH1* ORF was subcloned into the pET15b vector modified to contain a PreScission protease cleavage site. The vector containing the *A. annua* *ADH1* ORF was used to transform the *E. coli* BLR (DE3). Protein expression was induced by adding 0.4 mM isopropyl-β-D-thiogalactoside (IPTG) and cultures totalling 5 l were incubated at 16 °C for 16 h. *ADH1* was subjected to two rounds of Ni-column purification, analysis by SDS-PAGE and dialysis. The final fractions containing *ADH1* were pooled and dialysed against protein storage buffer (20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 200 mM KCl, 10% glycerol and 1 mM

dithiothreitol (DTT)). Protein concentration was determined by Bradford assay and aliquots were stored at -80 °C. *ADH1* purity by SDS-PAGE was judged to be 95%.

Unless otherwise stated, *ADH1* enzyme assays included 50 mM Tris buffer, pH 8.5, 250 mM NaCl, 0.4 mg ml⁻¹ BSA, 50 μM substrate, 1 mM NAD, 3 μg of octadecane (as internal standard; Sigma-Aldrich) and 80 ng of recombinant *A. annua* *ADH1* in a total volume of 200 μl. Negative controls were carried out in the absence of NAD. Reactions were allowed to proceed for 4 min at 30 °C with shaking (500 r.p.m.), and immediately stopped by extraction with 500 μl pentane. All quantitative analyses were done with 3–6 technical replicates per treatment. Pentane extracts were concentrated to ~30 μl under a stream of nitrogen and either 10 μl ethyl acetate or 10 μl of a mixture of 1:1 *N,O*-bis-(trimethylsilyl)acetamide (Sigma-Aldrich)/pyridine (Fluka) was added. The remainder of the pentane was carefully removed under a stream of nitrogen and the final 10-μl sample was analysed by gas chromatography-mass spectrometry (GC-MS)²⁶.

Substrate specificity was determined in 15-min, 600-μl assays using (+)-borneol (Fluka), (-)-borneol (Fluka) and artemisinic, dihydroartemisinic, artemisia²⁶, coniferyl (Sigma-Aldrich), and cinnamyl (Sigma-Aldrich) alcohols. *ADH1* only showed considerable dehydrogenase activity with artemisinic alcohol and to a lesser extent with dihydroartemisinic alcohol (4.2% relative to artemisinic alcohol; Supplementary Fig. 5). The identity of the aldehyde products was confirmed by GC-MS in comparison to authentic standards. When assayed with artemisinic alcohol and NAD, recombinant *A. annua* *ADH1* showed a pH optimum of 8.5. Using 1 mM NADP as the cofactor, oxidation of artemisinic alcohol by *ADH1* was 30-fold lower (Supplementary Fig. 5). The linear range of the *ADH1* assay with respect to time was tested by reactions under standard assay conditions except varying the time up to 30 min. The pH optimum of the purified *ADH1* was determined to be 8.5 based on a series of 15-min assays with the pH range from 5.5 to 10 in intervals of 0.5-pH units using 50 mM citrate, phosphate, Tris, CHES and CAPS buffers. Kinetic parameters were determined by varying the concentrations of artemisinic alcohol (3.0–25 μM) in 600-μl assays using 240 ng *ADH1*. Substrate solubility prevented the use of higher concentrations. Octadecane was used as an internal standard to quantify the substrate and product from the reactions by gas chromatography using response factors determined by using known concentrations of standards. Kinetic constants were determined by fitting the data to the Michaelis-Menten equation using nonlinear regression and EnzFitter software (Biosoft).

Media and growth conditions. Fermentation media: The media used for this work were based on media described previously²⁹. The trace metal solution contained 5.75 g l⁻¹ ZnSO₄·7H₂O, 0.32 g l⁻¹ MnCl₂·4H₂O, 0.47 g l⁻¹ CoCl₂·6H₂O, 0.48 g l⁻¹ Na₂MoO₄·2H₂O, 2.9 g l⁻¹ CaCl₂·2H₂O, 2.8 g l⁻¹ FeSO₄·7H₂O and 80 ml l⁻¹ 0.5 M EDTA, pH 8.0. The vitamin solution contained 0.05 g l⁻¹ biotin, 1 g l⁻¹ calcium pantothenate, 1 g l⁻¹ nicotinic acid, 25 g l⁻¹ myo-inositol, 1 g l⁻¹ thiamine HCl, 1 g l⁻¹ pyridoxal HCl and 0.2 g l⁻¹ *p*-aminobenzoic acid. The batch medium for all fermentations contained 19.5 g l⁻¹ glucose, 15 g l⁻¹ (NH₄)₂SO₄, 8 g l⁻¹ KH₂PO₄, 6.2 g l⁻¹ MgSO₄·7H₂O, 12 ml l⁻¹ vitamin solution and 10 ml l⁻¹ trace metal solution.

The batch medium also contained additional components depending on the strain and the process being run. For the glucose and ethanol mixed-feed process, for all strains except Y285, CuSO₄ was added to the batch medium to a concentration of 0.25 μM CuSO₄. For Y285, the batch medium contained 20 μM CuSO₄.

The bioreactor feed media also varied for the different processes and for different strains. All mixed glucose/ethanol processes used bioreactor feed base that contained 386 g l⁻¹ glucose, 9 g l⁻¹ KH₂PO₄, 5.12 g l⁻¹ MgSO₄·7H₂O, 3.5 g l⁻¹ K₂SO₄, 0.28 g l⁻¹ Na₂SO₄ and 237 ml l⁻¹ ethanol (95% v/v).

For the glucose and ethanol mixed feed process, two different feed media were prepared for the fermentation: pre-induction-feed media and induction-feed media (includes small molecule inducers and repressors). For both feed media, stock solutions of vitamins and trace metals were added to the bioreactor feed base as follows: 12 ml vitamin solution per litre feed base, and 10 ml trace metals solution per litre feed base. The pre-induction-feed media contained 0.25 μM CuSO₄ and no other additions. For fermentation of Y285, the pre-induction-feed media contained 20 μM CuSO₄.

To the induction-feed medium, two different inducers/repressors were added to the medium dependent on the strain. For Y285, concentrated solutions of galactose and methionine were added to the induction-feed medium to bring the final concentrations to 10 g l⁻¹ galactose and 1 g l⁻¹ methionine. For all other strains, concentrated solutions of galactose and CuSO₄ were added to the feed medium to bring the final concentrations to 10 g l⁻¹ galactose and 150 μM CuSO₄. All additions to the medium were made in a sterile hood.

Shake-flask media: Seed medium for pre-culture was the batch fermentation medium modified with the addition of 100 ml l⁻¹ succinate buffer (0.5 M, pH 5.0). For strain Y285, the concentration of CuSO₄ in the seed medium was 20 μM. For

all other strains tested (which contain a copper-repressible promoter controlling ERG9 expression), low-copper seed medium, containing only $0.25 \mu\text{M}$ CuSO_4 , was used.

Flask-production medium was modified seed media which contained 40 g l^{-1} glucose, 5 g l^{-1} galactose, 1.7 mM methionine and $150 \mu\text{M}$ CuSO_4 .

Shake-flask methods. To prepare seed vials, single isolates of each strain from agar plates were grown for 18–24 h in 20-ml low-copper seed medium containing $0.25 \mu\text{M}$ CuSO_4 . Cultures were then inoculated at an attenuation ($D_{600 \text{ nm}}$) of 0.05 into fresh low-copper seed medium and grown for a further 18–24 h to an $D_{600 \text{ nm}}$ of between 2 and 3 (measured using a Thermo Scientific Genesys 10 Vis spectrophotometer). Six-hundred microlitres of this culture was added to $400 \mu\text{l}$ of 50% glycerol and stored in 1-ml aliquots (20% glycerol (v/v) final) at -80°C .

To acclimate cells before inoculation into production medium, frozen seed vials were thawed to room temperature and inoculated into 20-ml low-copper seed medium. Cultures were grown for 18–24 h at 30°C with shaking at 200 r.p.m. The next day, the cultures were diluted to a $D_{600 \text{ nm}}$ of 0.05 in 20-ml low-copper seed medium and grown for ~18 h at 30°C with shaking at 200 r.p.m.

Cells from the second overnight acclimation were diluted to a $D_{600 \text{ nm}}$ of 0.05 in 250 ml unbaffled flasks containing 25 ml flask-production medium. Flasks contained an additional 5 ml IPM where indicated. All cultures were inoculated in triplicate and incubated at 30°C for 72 h with shaking at 200 r.p.m. in a humidified Innova incubator. Flasks were sampled periodically for growth ($D_{600 \text{ nm}}$), viability and product titres. Viability was measured using the LIVE/DEAD Funga Light yeast viability kit for flow cytometry (Invitrogen Corporation) and a Guava technologies EasyCyte Plus flow cytometer.

For the production of insoluble artemisinic acid in shake-flask cultures, 15.8 g l^{-1} of 95% ethanol was added to the production flask after 72, 96 and 120 h growth. The flask was inspected for formation of insoluble material after 144 h.

Glucose and ethanol mixed-feed process. Preparation of seed cultures and procedures for setting up and running glucose and ethanol mixed-feed fermentations have been described³. The production process was induced with the addition of 10 g l^{-1} galactose and 0.25 g l^{-1} methionine (for Y285), or 10 g l^{-1} galactose and $150 \mu\text{M}$ CuSO_4 (all other strains) to the bioreactor after the culture reached an $D_{600 \text{ nm}}$ value of approximately 50. At this time, the feed bottle containing pre-induction-feed medium was exchanged for a feed bottle containing induction-feed medium.

Mixed glucose/ethanol feed process with early induction/repression. The mixed glucose/ethanol feed process was modified by changing the time of induction (the time of addition of galactose and methionine, or galactose and high CuSO_4) from the time the culture reached a $D_{600 \text{ nm}}$ of 50 to the time of inoculation. The mixed glucose/ethanol feed process with early induction/repression was identical to the mixed feed process except that batch and feed media were modified. Immediately before inoculation, concentrated solutions of galactose, methionine and/or CuSO_4 were added to the batch medium to bring the final concentrations to 10 g l^{-1} galactose and 0.25 g l^{-1} methionine (for Y285), or 10 g l^{-1} galactose and $150 \mu\text{M}$ CuSO_4 (for all other strains). Only the induction-feed medium was used in the fed-batch phase of the process (no pre-induction medium). All other parameters were the same as the mixed glucose/ethanol feed process.

At later time points (42–96 h), in some fermentation runs of strains Y1283 and Y1284, artemisinic acid precipitated from the liquid fermentor broth. Solid precipitate was visible in the fermentor and adhered onto the side of the bioreactor and the head plate. During the runs, artemisinic acid concentration was still assayed from fermentor broth samples over the course of the fermentation as described below. However, for select fermentations of Y1284, artemisinic acid was also assayed at the end of the fermentation after complete solubilization of the precipitate by high pH treatment of the fermentor broth. At the end of the fermentation, the culture was adjusted to pH 8.1 with 10 M NH_4OH and allowed to stir at 1,500 r.p.m. (maximum rpm) at 30°C for at least 1 h to dissolve the precipitated artemisinic acid. After the pH adjustment, the cell broth was collected and water was added the tank to wash any residual precipitate from the tank. The water was adjusted to pH 9.1 with 10 M NH_4OH and allowed to stir at 1,200 r.p.m. overnight. Together, this provided a more accurate measurement of artemisinic acid at the final time point.

Mixed glucose/ethanol feed process with induction/repression and IPM. The addition of an IPM phase to yeast cultures was tested at the fermentor scale using strains Y1283 and Y1284. The fermentation process used was the mixed glucose/ethanol feed process with early induction/repression. The only process change was the addition of 200 ml IPM to the fermentor before inoculation. The initial aqueous batch volume of the fermentations was 0.7 l.

The IPM phase and the aqueous cell broth formed a well-mixed emulsion in the reactor at later times in the fermentation (>24 h). To assay artemisinic acid titre in

fermentations with IPM, samples of the combined IPM and cell broth mixture/emulsion were extracted with solvent. The mixture was first vortexed then added to the methanol/formic acid as described below. The concentrations measured by liquid chromatography with ultraviolet detection are reported in terms of grams per litre total volume (aqueous cell broth plus IPM). Using the ratio of aqueous cell broth volume to IPM volume at the time of sampling, the titres are converted to terms of grams per litre aqueous volume to allow for direct comparison with runs that do not use IPM.

At 30°C , artemisinic acid has a solubility of approximately $100\text{--}115 \text{ g l}^{-1}$ in IPM (empirically determined). At later times in the fed-batch fermentation, after a large volume of aqueous feed has been added to the fermentor, the ratio of IPM/ aqueous volume is significantly lower and the solubility limit could restrict additional production.

Ethanol pulse-feed process with IPM. Yeast strain Y1284 was tested in an ethanol pulse-feed process with the addition of IPM to the culture medium. The temperature, pH and dissolved oxygen were controlled at the set points described above. The batch medium for this process was the same as for the glucose/ethanol mixed-feed process with early induction, described above, except that no galactose was added (Y1284 does not require galactose for induction). Four-hundred millilitres of IPM was added to a starting aqueous fermentor volume of 0.8 l, before inoculation.

The feed for the ethanol pulse-feed fed-batch phase of the process was 95% (v/v) ethanol. Because none of the salts, trace metals or vitamins was soluble in 95% (v/v) ethanol, concentrated feed components were combined into a concentrated post-sterile addition (PSA) solution. The concentrated PSA solution consisted of 72.9 g l^{-1} KH_2PO_4 , 41.4 g l^{-1} $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 28.3 g l^{-1} K_2SO_4 , 2.3 g l^{-1} Na_2SO_4 , 1.2 mM CuSO_4 , 10 ml l^{-1} trace metals solution and 12 ml l^{-1} vitamin solution. The concentrated PSA solution was injected through a septum in the bioreactor head plate with a syringe once per day according to how much volume of 95% (v/v) ethanol volume had been delivered since the previous addition of feed components. One-hundred-and-twenty-four millilitres of concentrated PSA solution was added per litre of 95% (v/v) ethanol added.

After the batch carbon was consumed (detected as described above) the ethanol pulse-feed algorithm was initiated. As the culture grew and consumed O_2 , dissolved O_2 was maintained at 40% by an agitation cascade followed by oxygen enrichment (as described above). In the first phase of the fed-batch fermentation, before the stir rate of the reactor reached the maximum allowed for the unit, the pulse feed algorithm used stir rate (Stir) measurements to control ethanol feed delivery (Supplementary Fig. 14a). The computer algorithm assigned a variable (Stir Max) that tracked the maximum stir rate obtained so far in the process. While growing on ethanol, O_2 demand increased and stir rate increased until the substrate was depleted from the fermentor medium. At that point, the dissolved O_2 increased and the controller decreased stir rate to maintain dissolved $\text{O}_2 = 40\%$. When Stir decreased to less than 75% of the value of Stir Max, the ethanol feed pump was activated for the length of time necessary to add 10 g ethanol per litre fermentor volume to the reactor. The computer algorithm calculated the time necessary to add 10 g ethanol per litre fermentor volume to the reactor (Timer Max) after each cycle. The first phase of the algorithm iterated unit O_2 enrichment as required.

After the stir rate of the reactor reached the maximum allowed for the unit, oxygen enrichment was used to maintain dissolved $\text{O}_2 = 40\%$. During this stage of the fed-batch fermentation, the second phase of the control algorithm was initiated. Dissolved O_2 measurements were used to control ethanol feed delivery (Supplementary Fig. 14b). When ethanol was depleted from the fermentor medium, the dissolved O_2 began to increase rapidly—faster than the dissolved O_2 controller could compensate. When dissolved $\text{O}_2 > 50\%$, the ethanol feed pump was activated for the length of time necessary to add 10 g ethanol per litre fermentor volume to the reactor. After the addition of ethanol, the dissolved O_2 would rapidly decrease to $<50\%$. The variable Timer Max was again calculated by the computer algorithm after each cycle. This algorithm iterated for the remainder of the fermentation.

Purification of artemisinic acid from IPM. IPM was isolated from artemisinic acid fermentations by centrifugation. IPM was mixed with 1% $\text{NaH}_2\text{PO}_4 \cdot 12\text{H}_2\text{O}$ and the pH was adjusted to 10.7 by the addition of 5 M NaOH. The solution was then stirred at ambient temperature for 60 min. After mixing, the solution was allowed to separate by gravity in a separatory funnel at ambient temperature. The bottom aqueous phase was drawn off from the upper IPM phase. The bottom aqueous phase was run through a liquid: liquid annular centrifugal contactor (CINC Industries) to ensure complete removal of any residual IPM. A 10% (w/v) SDS solution was added to the aqueous phase to bring the final SDS concentration to 0.03%. The solution was mixed and the pH adjusted to 5.0 with 2.5 M H_2SO_4 . The acidification resulted in the formation of a fine white precipitate, which was captured on a $0.45\text{-}\mu\text{m}$ PTFE (polytetrafluoroethylene) filter, rinsed

with purified water and then dried. Analysis of the IPM before and after aqueous extraction showed that 5% of the artemisinic acid remained in the IPM after extraction (~95% step yield). Analysis of the filtrate after precipitation showed that 2% of the artemisinic acid present in the aqueous phase remained in the filtrate after acidification (~98% step yield). The overall purification yield obtained was ~93%. Additional aqueous extractions of the remaining IPM should increase the overall yield. Analysis of the dried precipitate by gas chromatography with flame-ionization detection (GC-FID) gave artemisinic acid purities of ~96% by area and ~98% by weight.

Broth extraction. Amorpha-4,11-diene, artemisinic alcohol and artemisinic aldehyde were extracted from cells and broth as follows. Cell lysis cocktail was prepared by combining two parts Novagen YeastBuster protein reagent (EMD Biosciences) and one part 2 M HCl. Samples were prepared by mixing 0.4 ml cell lysis cocktail with 0.1 ml whole broth and 1 ml ethyl acetate containing 10 mg l⁻¹ *trans*-caryophyllene (internal standard, ≥98.5% purity; Sigma-Aldrich) in a 2-ml glass vial. The sample was mixed for 30 min on a vortex mixer. After mixing, the vial was placed on the bench top to allow the phases to separate. If necessary, the vial was centrifuged at 1,000g to break any emulsion that had formed. Six-hundred microlitres of the ethyl acetate layer was transferred to a gas chromatography vial for analysis.

Gas chromatography. The production of amorpha-4,11-diene, artemisinic alcohol and artemisinic aldehyde was monitored by GC-FID. The ethyl acetate-extracted samples were analysed using on the GC-FID. Amorpha-4,11-diene, artemisinic alcohol and A.CHO peak areas were converted to concentration measurements from external standard calibrations using authentic compounds. To expedite run times, the temperature program and column were modified to achieve optimal resolution and the shortest overall run-time with minimal interferences. A 10-μl sample was split 1:20 and was separated using a DB-WAX column (50 m × 200 μm × 0.2 μm; Agilent), with hydrogen as the carrier gas at a flow rate of 1.57 ml min⁻¹. The temperature program for the analysis was as follows: the column was initially held at 150 °C for 3 min, followed by a temperature gradient of 5 °C min⁻¹ to a temperature of 250 °C, and then the column was held at 250 °C for 5 min to elute all remaining components. Under these conditions, *trans*-caryophyllene, amorpha-4,11-diene, artemisinic aldehyde and artemisinic alcohol elute at 4.95, 5.77, 12.94 and 18.60 min, respectively.

Broth preparation with and without IPM. A 1-ml aliquot of well-mixed fermentation broth was diluted in 9 ml of methanol plus 0.1% formic acid (IPM formed an emulsion with the cell broth when it was used). The mixture was then

mixed on a vortex mixer for 30 min and centrifuged at 16,000g for 5 min. One-hundred microlitres of the supernatant was diluted into 900 μl methanol plus 0.1% formic acid, and analysed by the HPLC method described below.

In-process assay for titre measurement. A screening method was developed to rank artemisinic-acid-producing strains. This method was used only to rank strains, and not determine final titre. A 20-μl aliquot was injected on an Agilent 1200 HPLC with ultraviolet detection at 212 nm. An Supelco Discovery C₈ column (4.6 mm × 100 mm × 5.0 μm; Supelco) equipped with the appropriate guard column (4.0 mm × 20.0 mm; Supelco) was used for separation, with the following gradient at a flow rate of 1 ml min⁻¹ (channel A: water plus 0.1% formic acid; channel B: methanol plus 0.1% formic acid): 0–0.5 min 70% B, gradually increased to 97% B from 0.5 to 6.7 min, held at 97% B until 7 min, decreased to 70% B from 7 to 7.5 min, and re-equilibrated to 70% B from 7.5 to 9.5 min. The column was held at 25 °C during the separation. Under these conditions, artemisinic acid was found to elute at 6.3 min. AA peak areas were converted to concentrations from external standard calibrations of authentic compounds.

Final titre measurement. A 20-μl aliquot was injected on an Agilent 1200 HPLC with ultraviolet detection at 212 nm. An Agilent Eclipse XDB-C₁₈ column (4.6 mm × 100 mm × 3.5 μm, Agilent) equipped with the appropriate guard column (4.6 mm × 12.5 mm × 5 μm, Agilent) was used for separation, with the following gradient at a flow rate of 1 ml min⁻¹ (channel A: water plus 0.1% formic acid; channel B: acetonitrile plus 0.1% formic acid): 0–10 min 50% B, gradually increased to 100% B from 10 to 21 min, held at 100% B until 26 min, decreased to 50% B from 26 to 26.1 min, and re-equilibrated to 50% B from 26.1 to 31 min. The column was held at 45 °C during the separation. Under these conditions, artemisinic acid was found to elute at 15.54 min. AA peak areas were converted to concentrations from external standard calibrations of authentic compounds.

26. Polichuk, D. R., Zhang, Y., Reed, D. W., Schmidt, J. F. & Covello, P. S. A glandular trichome-specific monoterpene alcohol dehydrogenase from *Artemisia annua*. *Phytochemistry* **71**, 1264–1269 (2010).
27. Zhang, Y. *et al.* The molecular cloning of artemisinic aldehyde Δ¹¹(13) reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in *Artemisia annua*. *J. Biol. Chem.* **283**, 21501–21508 (2008).
28. Graham, I. A. *et al.* The genetic map of *Artemisia annua* L. identifies loci affecting yield of the antimalarial drug artemisinin. *Science* **327**, 328–331 (2010).
29. van Hoek, P., de Hulster, E., van Dijken, J. P. & Pronk, J. T. Fermentative capacity in high-cell-density fed-batch cultures of baker's yeast. *Biotechnol. Bioeng.* **68**, 517–523 (2000).

CAREERS

NIH President's budget proposes 7% pay rise for entry-level postdocs **p.539**

FACULTY Report finds number of US non-tenure-track posts is growing **p.539**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



She is convinced that artists and designers stoke scientists' creativity.

Scientists are becoming increasingly open to artistic collaborations, which offer career benefits including improved productivity as a result of a new perspective or a more creative outlook; bolstered communication and outreach skills; and contacts among artists, like-minded scientists or funding agencies. Learning how to indulge artistic pursuits — and avoid professional obstacles such as being perceived as unfocused or undisciplined — is key to shaping a career that can sustain both art and science.

SUPPORT ON SHOW

Hybrid art-science efforts have gained support in recent years. Some institutions see them as a means of enhancing creativity and innovation, and a growing number are creating cross-disciplinary centres. Examples include the Media Lab at the Massachusetts Institute of Technology (MIT) in Cambridge and the Art|Sci Center + Lab at UCLA. "We are absolutely on the brink of a new renaissance," says James Gimzewski, a nanobiologist at UCLA who began collaborating with artists ten years ago in the hope of engaging and educating the public. Artistic collaborations seem to thrive particularly in newer areas of scientific exploration, including synthetic biology, nanotechnology, robotics and neuroscience.

"We're trying to raise the visibility of our interest in supporting art-science collaborative projects," says Bill O'Brien, senior adviser for programme innovation at the US National Endowment for the Arts (NEA) in Washington DC, which is increasingly directing funds to science- and technology-focused arts projects — responding, in part, to growing interest. It spent about US\$963,000 on such grants in 2012, up from \$304,000 in 2009.

Other major science funders are also fostering academic efforts to create art-science collaborations. Guna Nadarajan, dean of the University of Michigan School of Art & Design in Ann Arbor, is helping to build the NSF-funded Network for Sciences, Engineering, Arts and Design (SEAD) to help artists and scientists to connect and collaborate, and to explore how to conduct research at the intersection of art, science and engineering. So far, SEAD has 300 participants across 30 research institutions and art colleges.

Meroë Candy, senior arts adviser to the Wellcome Trust in London, one of the world's largest biomedical-research funders, says ►

Using crystal blocks, an exhibit displays the flights of bees in three dimensions as they learned a task.

INTERDISCIPLINARITY

Artistic merit

Options abound for scientists who want to get in touch with their inner artist, whether professionally or as a hobby.

BY VIRGINIA GEWIN

Christina Agapakis thought that her interest in art would always be separate from her pursuit of science, especially once she had decided to do postgraduate studies in biology.

But she found ways to meld the two. While earning her PhD in biomedical sciences at Harvard University in Cambridge, Massachusetts, Agapakis joined a social experiment called Synthetic Aesthetics. A joint project of the University of Edinburgh, UK, and Stanford University in California, funded by the US National Science Foundation (NSF) and the UK Engineering and Physical Sciences Research Council, Synthetic Aesthetics teamed artists and designers with synthetic biologists and encouraged them to come up with

interdisciplinary ideas and projects. Agapakis worked with Sissel Tolaas, a researcher and scent artist resident in Berlin. Together, they made cheese — using starter cultures made of bacteria isolated from the human body. They wanted to make the unseen biological world perceptible to the senses, and to call attention to how synthetic biology might alter microbial communities. "The creativity of designing, rather than studying, biology is really exciting for me," says Agapakis.

Now a postdoc in synthetic biology at the University of California, Los Angeles (UCLA), Agapakis continues to collaborate with designers, incorporating principles of balance and scale — in this case, being mindful of microbial relationships and interactions — into the design of microbial communities that could, for example, yield new fertilizers or biofuels.

► that the organization's arts budget has grown from £100,000 (US\$153,000) in 1996 to £1.4 million this year.

Industry, too, has discovered the potential of artistic aspirations. "Executives are eagerly hiring people who bring a key element of creativity to produce game-changing ideas," says Nadarajan.

Artistic interests often help scientists to enhance their own creativity in the lab. After 11 years as an evolutionary biologist at the University of Montreal in Canada, François-Joseph Lapointe was restless — so in 2005, he started a second PhD in dance. These days, he pursues both dance and scientific research. When his science focused on finding genetic signals of evolutionary lineages, for example, he developed choreography that assigned movements to each DNA nucleotide, and performers danced out their own genetic codes. He has begun work on metagenomics, or the study of genetic material in a particular environment, and hopes to sequence his dancers' microbial genomes. Some colleagues suggest that splitting his time means that he is shortchanging his science, but he disagrees. "I am happier and more productive when I use my brain differently," he says.

Meaningful scientific advances can benefit from an artistic perspective, says Gimzewski. Scientists often think reductively, in terms of phenomena isolated from their environment; artists, by contrast, observe and study inter-related phenomena and then craft an interpretation. For three years, Gimzewski has been working on a project to build an artificial brain, funded by the US Defense Advanced Research Projects Agency, and he says that he would never have tackled such a complex project without his visual-arts experience, which changed his science. "I used to look at single molecules, but it's essential in the world today to work in complex environments," says Gimzewski.

TWO-WAY PARTNERSHIP

In the past, art-science collaborations have tended to begin when artists hoping to learn about science have approached researchers. But increasingly, scientists are being proactive, seeking artistic tools to bolster their research. Karissa Sanbonmatsu, a molecular biologist at Los Alamos National Laboratory (LANL) in New Mexico, partnered with specialists in three-dimensional (3D) visualization at the LANL's supercomputing facility to make sense of ribosomes, cell organelles that can have several hundred thousand atoms in one molecule. "We couldn't see the big picture using straightforward computer code, but the 3D images made troubleshooting much easier," she says. Her collaborations have changed how she thinks about her research, inspiring new directions — in one example, seeing the ribosome's apparently random movements made her curious about how their gyrations affect how well they function. Now Sanbonmatsu has joined more than two dozen other scientists



François-Joseph Lapointe choreographs dances based on DNA sequences.

in the Scientists/Artists Research Collaborations initiative at the Santa Fe Institute in New Mexico, which aims to focus on issues of climate change and energy.

When artist and scientist come together, it is important to find the best working arrangements for both parties, and for the project. All collaborators should articulate their goals and intentions clearly. "Having artists in residence in my lab just didn't work for me," says Beau Lotto, a neurobiologist at University College London. "I didn't find it interesting to have them gleaning ideas only to go away and make art." So Lotto created a studio space in which artists and scientists could interact and conduct research together on perception and human behaviour. He even started running a monthly night club to observe people in a real-world setting. "The only way you get questions that have never really been asked before is to bring in different perspectives and question assumptions," he says.

Scientists may find that tapping into the art community is a good way to make contacts and raise their own profiles. "I have access to twice as many grants, potential students and conferences," says Lapointe. This year, he has submitted grant proposals to the Natural Sciences and Engineering Research Council of Canada; the Social Sciences and Humanities Research Council of Canada; and the Canada Council for the Arts.

Lotto found that funders often support the marriage of art and science to engage the public. But he no longer fell under that remit once he started conducting research, and he lost funding. He has since launched his own fund-raising efforts, and has made money from donations at art installations and corporate sponsorship of talks on creativity, among other ventures.

Some funders are experimenting. The European Union has granted €1.6 million (US\$2.1 million) to StudioLab, a Europe-wide consortium of arts and science centres that helps scientists and artists to create interactive outreach events looking at the future of water resources, synthetic biology and the future of

social interactions. One project, *Biohacking: Do It Yourself!*, which ran from January to March at the Medical Museion in Copenhagen, advises viewers on how to conduct scientific experiments at home using moderately priced equipment.

The Hub, launched by the Wellcome Trust in January, offers £1 million and a space at the Wellcome Collection museum and art gallery in London to teams conducting interdisciplinary art and science research. Applications are due by 3 May.

ARTFUL STUDENTS

Graduate students who want to pursue both art and science have multiple options. Universities are responding to the demand from students with hybrid interests, who want to pursue the coupling of art and science rather than be forced to choose between the two, says Roger Malina, an astronomer at the University of Texas at Dallas and editor-in-chief of *Leonardo*, the journal of the International Society for the Arts, Science and Technology. His university opened an Arts and Technology (ATEC) programme in 2004; last year, Malina, launched an ATEC PhD programme that currently has 55 students and is planning to double in size in the next few years. "I'm sceptical of hyped-up claims that art-science is the next big thing, but I think it's really important and will definitely keep growing," says Malina.

He adds that other universities are also experimenting with how best to fuse art and science. Some, such as MIT, UCLA and the University of California, Davis, offer student training at their art-science centres or labs. In France, a partnership across the scientific research institutes and the decorative- and performing-arts centres of Paris Science and Letters has launched the Science, Art, Creation, Research PhD programme. Students are, for example, creating living pictures with microalgae.

And such training can help newly minted PhD holders to expand their job search to



Christina Agapakis has created cheese using starter cultures from the human body.

SENG HOK NGO

NICK SEEVER

VICTORIA VESNA

include art-related posts. “We are starting to see a few positions for hybrid art–science professionals, and I believe this will continue to grow,” says Malina. Nadarajan notes that Google and IBM are hiring graduates with design backgrounds for their research and development teams; and companies such as 3M and Proctor & Gamble have a steady demand for those skills in their efforts to develop innovative materials.

Neuroscientist Siddharth Ramakrishnan is convinced that his work — which included an interactive exhibit focused on the *Hox* genes that define body regions in all animals — proved beneficial during his job search last year. “My art collaborations helped me stand out rather than being just one of hundreds of other neuroscientists who had done successful postdocs,” he says, noting that interviewers found his proposals for campus-based art–science salons intriguing. In October, he started a job as an assistant professor at the University of Puget Sound, a liberal-arts college in Tacoma, Washington.

UNCERTAIN PATH

Not all research institutions or scientist colleagues embrace art–science collaborations. “You have to be aware that you could possibly jeopardize your career,” says Lotto. “Many universities don’t know how to assess the output of collaborations and some even actively discourage them.”

Steve Potter, a neuroscientist at Emory University in Atlanta, Georgia, agrees. For the past decade, he has worked with artists at SymbioticA, an art–science studio at the University of Western Australia in Perth. One project, called MEART, connected a robotic arm to a network of rat neurons cultured on a multi-electrode array, to study the essence of creativity. But even though his department supports his endeavours, Potter knows that some colleagues are less accepting. He advises young scientists eager to pursue dual interests to consider joining an art department. “The safest thing to do is join a department that is open-minded; often that is more likely to be an art department,” he says.

Systems engineer Leila Madrone says that aspiring artist–scientists should not despair if they have to do their art on the side at first. “Sometimes my work and art interests merge and sometimes they separate. It’s most important to work in a creative environment,” she says. As an undergraduate in MIT’s media lab, Madrone combined giant Tesla coils, which put out stunning arcs of high-voltage electricity, with robots to create an interactive musical performance. In 2006, she headed to NASA’s Ames Research Center in Moffett Field, California, to join the Intelligent Robotics Group. There, she worked on GigaPan, a robotic panoramic image-capture system. Now she is one of about 25 engineers at Otherlab, an independent engineering lab



The ‘Blue Morph’ uses images and sounds from the metamorphosis of a caterpillar to a butterfly.

in San Francisco, California, that focuses on innovation in areas such as robotics, solar energy and electric vehicles. She does not have the security and benefits that she might get at a larger, established company, but there are perks. “I get to define what I am doing — which is why, I think, people are attracted to this path,” she says.

That path is not for everybody. “There is no recipe for a career in art–science,” says Malina. Rather than looking for a formula or a well-trodden path, he says, students should identify specific career goals and develop the skills to achieve them, such as learning computer programming and design principles. And students might consider whether those hybrid skills are best suited to distinguish their art–science research aims, attract collaborators or simply provide a vehicle for artistic expression.

Agapakis is confident that she will continue to create her own opportunities. She has just finished a three-week stint helping to teach a graduate-level media design course focused on biotechnology at the Art Center College of Design in Pasadena, California. “For me,” says Agapakis, “playing it safe is riskier because I wouldn’t be pursuing the things I’m most passionate about.” ■

Virginia Gewin is a writer based in Portland, Oregon.

CORRECTION

The Careers Brief ‘Online journal club’ (*Nature* **496**, 261; 2013) wrongly gave the impression that the journal club mentioned was the first to go online; it was, in fact, the first to use the Journal Club Live platform.

NIH

Postdoc pay rise

Entry-level postdocs funded by the US National Institutes of Health would get a 7% stipend increase next year under President Barack Obama’s proposed 2014 federal budget. New PhD recipients who receive the Ruth L. Kirschstein National Research Service Award (NRSA) would earn US\$42,000; those with a year or more of experience would receive a 4% rise over existing levels. “This is a huge step forward in recognizing the value of postdoctoral researchers’ contributions,” says Cathée Johnson Phillips, executive director of the National Postdoctoral Association (NPA) in Washington DC, which since 2001 has been advocating for the entry-level stipend to increase to \$45,000. The stipend rose by 1% in 2009 and 2010, and by 2% in 2011 and 2012. A 2011 NPA survey found that half of US institutions base postdoctoral pay on the NRSA.

FACULTY

Non-tenured jobs grow

The number of full-time non-tenure-track faculty members at US institutions grew by about 13% from 2007 to 2011, compared with 11% for part-time faculty members, a report finds. *Here’s the News: The Annual Report on the Economic Status of the Profession, 2012–13*, published on 8 April by the American Association of University Professors (AAUP) in Washington DC, also notes that more than one-fifth of assistant professors were off the tenure track in 2010–11. “Even among ranks that we would think of as tenure track, a significant proportion of faculty are not,” says John Curtis, AAUP director of research and public policy.

UNITED STATES

Chinese applications fall

The number of applications to US graduate schools from students in China fell this year for the first time since 2006, when the US Council of Graduate Schools (CGS) in Washington DC first had sufficient sample sizes to keep track. An 8 April CGS report finds that total international applications to the United States grew by 1%, the smallest rise in 8 years. Chinese applications fell by 5%, after growth of 19% in 2012 and 21% in 2011. Services that help students to narrow down their choices may be a factor, says Rajika Bhandari, deputy vice-president for research and evaluation at the Institute of International Education in New York, as may the cost of US applications.

THE EPISTOLARY HISTORY

Time for a change.

BY ALEX SHVARTSMAN

1: 1/9/12

Hey Cat,

We finally did it! The time machine works. The blokes are talking about trying to sell it to some big technology company, but I have a better idea.

A quick and easy trip to grand-grand-grandpa Oskar's machine shop in 1890 Weimar, a couple of sketches and a sample left on his desk, and presto: Oskar invents duct tape and builds a fortune in Germany; enough of it gets passed on to my branch of the family a century later that we don't need any vulture capitalists grabbing the lion's share of the time-travel tech profits. Besides, with a little one on the way we can use the extra dough.

So I'm e-mailing to let you know that I'm staying at Oxford to work on this tonight and might miss dinner. On the bright side, if things work out how I expect them to, we'll be dining on caviar instead of pizza.

2: September 01, 2012

My Dear Cathy,

Yesterday was the happiest day of my life. I finally perfected my invention, but the news of your pregnancy is a miracle that outshines any achievements of mere science.

I couldn't sleep last night, thinking of the world our son or daughter will be born into. England ravaged by 70 years of total war and the constant Nazi air raids — it's not the sort of place in which I want them to spend their childhood.

With a working prototype of the time machine in hand, I have both the means and the moral responsibility to fix the mistakes of the past. I'm going to travel back to 1930 and kill Hitler.

If all goes well, you'll wake up and read this note in a far better world.

3: Сентябрь 01, 2012

Dear Katya,

My comrades at the Oxford Universitet and I have finally perfected the device. We're scheduled to present Project 'Machina Vremeny' to the Politburo in the morning. When you shared

the great news last night, I couldn't sleep, thinking of the world our children will be born into. I can't stand the thought of them living in the constant fear of nuclear annihilation that is hanging over all the free people of Socialist Europe.



I possess the means and the moral authority to prevent 70 years of the cold war. I'm going to travel back to 1930 and kill Roosevelt.

If all goes well, you'll wake up and read this note in the better world, one where communism has already been achieved.

7: First day of September in the year of our Lord two thousand and twelve

Dearest Catherine,

I received your kind letter a few days since and am dreadfully sorry that the fertility infusions are not yet working. I direct this letter to you in hope that my own fortuitous developments shall cheer your heart and improve your disposition.

The Chronomat device I've endeavoured to design is finally complete. My lifelong dream of single-handedly defending Her Majesty's Empire against those belligerent ruffians from the American colonies is within my grasp. Two centuries of combatting the rebels have sapped our resources and

surely delayed technological process. By God, we don't even yet have the steam-powered flying carriage, the invention of which the fictioneers of old predicted to occur back in the 1970s.

The world would have been a better place had the civilized man never ventured into the Americas, and thence I shall presently activate the Chronomat and use it to prevent Mr Columbus from undertaking his journey.

By the time this letter reaches you at the clinic, we shall all be living in a better tomorrow.

14: Haab: 12 Mol. Tzolkin: 10 Muluc

Dear Diary,

Once again, I failed to meet a suitable partner today.

I dragged myself to the drinking hall, but there were few single women there, and none of them interested in my advances. Instead, I found myself drinking alone and listening to a pair of inebriated Maya who were apparently anxious about an impending end of the world.

Their main argument seemed to be that the ancient Christian calendar extended no farther than 2012. As if the priests of an extinct Eurasian cult possessed the scientific knowledge to predict some future catastrophe. Absurd!

I went home, alone. I couldn't sleep, lying in bed and imagining what it might be like to invent the means of changing the past. How different would our world be if the Mayan explorers had never arrived at the shores of Europe all those centuries ago? What sort of culture and science could the pale-faced tribes of this continent have developed if they weren't wiped out or subjugated by the superior Western civilization?

We'll never know. Travelling back in time is a silly fantasy I conceived of only due to imbibing too much balché yesterday evening.

I shall purge such thoughts from my mind, bathe, rest and prepare myself. Tomorrow, I shall go out and try again. Somewhere out there is a woman who is destined to be my soul mate. I haven't met her yet, but I remain an optimist. ■

Alex Shvartsman is a writer and game designer from Brooklyn, New York. His other fiction is linked at www.alexshvartsman.com.

JACEY

➔ NATURE.COM

Follow Futures:

Twitter @NatureFutures

Facebook go.nature.com/mtoodm